DOCUMENT RESUME

ED 243 407 HE 017 318

TITLE	Quality in the Pell Grant Delivery System. Volume 3: Procedures and Methods. Final Report.
INSTITUTION	Advanced Technology, Inc., Reston, VA.
SPONS AGENCY	Office of Student Financial Assistance (ED), Washington, DC.
PUB DATE	30 Apr 84
CONTRACT	300-80-0952
NOTE	169p.; For related documents, see ED 217 788-791. For Executive Summary, see HE 017 315; for Volumes 1 and 2, see HE 017 316-317.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Statistical Data (110)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Compliance (Legal); *Delivery Systems; Error Patterns; Federal Aid; Federal Programs; Financial Aid Applicants; *Grants; Higher Education; National Surveys; Program Validation; *Quality Control; *Research Methodology; Resource Allocation; Sampling; *Student Financial Aid
IDENTIFIERS	*Pell Grant Program; Response Rates (Questionnaires)

ABSTRACT

The Pell Grant Quality Control Study of 1982-83 was designed to identify program error rate, to measure the impact of increased validation activity, and to propose corrective actions to reduce the misallocation of program funds. A sample of approximately 4,000 students was drawn from a stratified random sample of 317 participating institutions. The results shoed that Pell Grant recipients in 1982-83 were granted 13 percent more than they should have been, although both student and institutional error dropped from 1980-81. The study confirmed that institutions complied with the revised validation requirements for the Pell Grant program in 1982-83. Volume 3 of this study reports on the procedures and methods used in the investigation. It includes: (1) sample section (sampling plan, sampling frame, selection of students from sample institutions); (2) institutional visits (data collectors, scheduling, training, field supervision, etc.); (3) student and parent data collection (supervisor and interviewer training, field operations, the Automated Survey Control System, secondary data collection); (4) data processing (institutional data, individual data); (5) data analysis (best value selection, detailed research questions, statistical analysis, control groups); and (6) survey response rates (institutional response rates, student and parent survey responses, variance estimates). (Author/LB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243407

FINAL REPORT

QUALITY IN THE PELL GRANT DELIVERY SYSTEM

VOLUME 3

PROCEDURES AND METHODS

Submitted to

OFFICE OF STUDENT FINANCIAL ASSISTANCE
DEPARTMENT OF EDUCATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CONTRACT NO. 300-80-0952

MARCH 1984

ADVANCED TECHNOLOGY, INC.
12001 Sunrise Valley Drive
Reston, Virginia 22091

and

WESTAT, INC.
1650 Research Boulevard
Rockville, Maryland 20850

TABLE OF CONTENTS

Chapter			Page
LIST OF FIGURES			v
LIST OF TABLES			vi
1	INTRODUCTION		1-1
2	SAMPLE SECTION		2-1
	2.1	Overview of Sampling for the Study	2-1
		2.1.1 The RDS Sample	2-1
		2.1.2 The ADS Sample	2-3
		2.1.3 The Control Group Samples	2-3
	2.2	Detailed Sampling Plan	2-4
		2.2.1 Sample Size and Precision	2-4
		2.2.2 Design Objectives	2-6
	2.3	Sampling Frame: Institutions Using Regular Delivery System	2-6
		2.3.1 Selection of a Single Report Per Institution	2-6
		2.3.2 Selection of Eligible Institutions	2-7
		2.3.3 Imputation of Recipients (Measure of Size)	2-7
		2.3.4 Treatment of Campuses Where the Central Office Reports for all Branches	2-8
		2.3.5 Determination of Certainty Institutions	2-8
		2.3.6 Selection of Noncertainty Institutions	2-9
	2.4	Selection of Students from Sample RDS Institutions	2-10
		2.4.1 Sampling Frame	2-12
		2.4.2 Weighted Estimates	2-13
3	INSTITUTIONAL VISITS		3-1
	3.1	Data Collectors	3-1
		3.1.1 Recruitment	3-2
		3.1.2 Qualifications	3-2
	3.2	Scheduling	3-3
		3.2.1 Master Schedule	3-3
		3.2.2 Schedule Confirmation	3-4

TABLE OF CONTENTS (Continued)

Chapter			Page
3.3	Training		3-6
	3.3.1	Manuals	3-18
	3.3.2	Interviewer Training	3-18
	3.3.3	Student Record Abstract Training	3-19
	3.3.4	Site Visit Practice	3-19
3.4	Field Supervision		3-20
	3.4.1	Scheduled Calls	3-20
	3.4.2	Technical Liaison	3-20
	3.4.3	Monitoring	3-21
	3.4.4	Validation	3-21
	3.4.5	Problem Resolution	3-24
	3.4.6	Financial Reporting and Cost Management	3-24
	3.4.7	Schedule Changes	3-25
3.5	Expansion of Control Group Data		3-25
3.6	Field Personnel Problems		3-27
3.7	Debriefing		3-28
4	STUDENT AND PARENT DATA COLLECTION		4-1
4.1	Organization of the Field Force		4-1
4.2	Supervisor and Interviewer Training		4-3
4.3	Field Operations		4-4
4.4	The Automated Survey Control System (ASCS)		4-6
4.5	Secondary Data Collection		4-6
	4.5.1	IRS Tax Returns	4-7
	4.5.2	Financial Institution Records (FIR)	4-8
	4.5.3	Tax Assessor Records (TAR)	4-8
5	DATA PROCESSING		5-1
5.1	Introduction		5-1
5.2	Institutional Data		5-5
	5.2.1	Data Control Procedures	5-5
	5.2.2	Coding Procedures	5-10
	5.2.3	Data Entry Procedures	5-12
	5.2.4	SAS Merge Procedures	5-14
	5.2.5	Documentation	5-18
	5.2.6	Statistical Analysis	5-18
	5.2.7	Software Quality Control	5-20

TABLE OF CONTENTS (Continued)

Chapter			Page
5.3	Individual Data		5-22
	5.3.1	Receipt Control	5-23
	5.3.2	Coding	5-23
	5.3.3	Data Retrieval	5-25
	5.3.4	Key Entry and Machine Editing	5-25
6	DATA ANALYSIS		6-1
6.1	Best Value Selection		6-1
	6.1.1	Strength of Data Sources and Documentation	6-2
	6.1.2	Premises in Best Value Selection	6-9
6.2	Detailed Research Questions		6-12
6.3	Statistical Analysis and Table Production		6-13
6.4	Control Groups		6-13
7	SURVEY RESPONSE RATES		7-1
7.1	Institutional Survey Response Rates		7-1
7.2	Student and Parent Survey Response		7-2
	7.2.1	Survey Response Rate	7-2
	7.2.2	Sample Loss	7-3
	7.2.3	Response Rate for Analytical Purposes	7-5
7.3	Variance Estimates		7-6
	7.3.1	Variance Estimation by the Method of Balanced Repeated Replication	7-6
	7.3.2	Estimated Sampling Errors	7-8

LIST OF FIGURES

Figure		Page
3-1	Data Collection Regions	3-5
3-2	Site Visit Scheduling Script	3-7
3-3	Training Schedule	3-13
3-4	Interviewer Validation Report	3-22
3-5	Field Memorandum on Control Group Data Collection	3-26
3-6	Debriefing Agenda	3-29
4-1	Westat Superivsory Regions for Student-Parent Interviewing	4-2
5-1	Pell Grant Quality Control Project--Stage Three Data Processing Plan Data Flow	5-3
5-2	Pell Grant Quality Control Project--Merge and Analysis	5-4
5-3	Student Listing Log	5-6
5-4	Institutional Control Sheet	5-7
5-5	Batch Control Sheet	5-8
5-6	Data Base Merge and Best Value Selecting Processes	5-16
5-7	Dataset Documentation Form	5-19
5-8	Program Development	5-21
6-1	Best Value Selection Priorities: Adjusted Gross Income for Independent Students	6-3
6-2	Best Value Selection Priorities: Home Value for Parent(s) of Dependent Students	6-6
6-3	Best Value Selection Priorities: Support by Parents 1981 for All Students	6-8
6-4	List of Items with Best Value Selection Priorities	6-10

LIST OF TABLES

Table		Page
2-1	Expected Coefficient of Variation for Average Absolute Dollar Value of Error, Assuming Different Values of P	2-6
2-2	Distribution of Recipients in Institution Universe Compared to Sampled Institutions	2-11
7-1	Preliminary Student/Parent Survey Reponse Rates	7-4
7-2	Disposition of Student and Parent Survey Sample Loss Cases	7-5
7-3	Estimated Sampling Errors: Selected Program-Wide Dollar Estimates	7-9
7-4	Estimated Sampling Errors: Selected Number of Recipients with Error	7-10
7-5	Estimated Sampling Errors: Selected Mean Dollar Estimates	7-12
7-6	Estimated Sampling Errors By Selected Categories: Institutional Overaward	7-13
7-7	Estimated Sampling Errors by Selected Categories: Student Overaward	7-14
7-8	Estimated Sampling Errors by Selected Categories: Case Overaward	7-15
7-9	Estimated Sampling Errors by Selected Cartegories: Institutional Underaward	7-16
7-10	Estimated Sampling Errors by Selected Categories: Student Underaward	7-17
7-11	Estimated Sampling Errors by Selected Categories: Case Underaward	7-18
7-12	Estimated Sampling Errors by Selected Categories: Net Institutional Error	7-19
7-13	Estimated Sampling Errors by Selected Categories: Absolute Institutional Error	7-20
7-14	Estimated Sampling Errors by Selected Categories: Net Student Error	7-21
7-15	Estimated Sampling Errors by Selected Categories: Absolute Student Error	7-22

LIST OF TABLES (Continued)

Table		Page
7-16	Estimated Sampling Errors by Selected Categories: Net Case Error	7-23
7-17	Estimated Sampling Errors by Selected Categories: Absolute Case Error	7-24
7-18	Estimated Sampling Errors by Selected Categories: Number of Recipients with Institutional Overaward	7-25
7-19	Estimated Sampling Errors by Selected Categories: Number of Recipients with Institutional Underaward	7-26
7-20	Estimated Sampling Errors by Selected Categories: Number of Recipients with Student Overaward	7-27
7-21	Estimated Sampling Errors by Selected Categories: Number of Recipients with Student Underaward	7-28
7-22	Estimated Sampling Errors by Selected Categories: Number of Recipients with Case Overaward	7-29
7-23	Estimated Sampling Errors by Selected Categories: Number of Recipients with Case Underaward	7-30
7-24	Estimated Sampling Errors by Selected Categories: Mean Institutional Overaward	7-31
7-25	Estimated Sampling Errors by Selected Categories: Mean Student Overaward	7-32
7-26	Estimated Sampling Errors by Selected Categories: Mean Case Overaward	7-33
7-27	Estimated Sampling Errors by Selected Categories: Mean Institutional Underaward	7-34
7-28	Estimated Sampling Errors by Selected Categories: Mean Student Underaward	7-35
7-29	Estimated Sampling Errors by Selected Categories: Mean Case Underaward	7-36
7-30	Estimated Sampling Errors by Selected Categories: Mean Net Institutional Error	7-37
7-31	Estimated Sampling Errors by Selected Categories: Mean Absolute Institutional Error	7-38

LIST OF TABLES (Continued)

Table		Page
7-32	Estimated Sampling Errors by Selected Categories: Mean Net Student Error	7-39
7-33	Estimated Sampling Errors by Selected Categories: Mean Absolute Student Error	7-40
7-34	Estimated Sampling Errors by Selected Categories: Mean Net Case Error	7-41
7-35	Estimated Sampling Errors by Selected Categories: Mean Absolute Case Error	7-42
7-36	Estimated Sampling Errors by Selected Categories: Number of Recipients with Unsatisfactory Academic Progress Eligibility Error	7-43
7-37	Estimated Sampling Errors by Selected Categories: Number of Recipients with Less Than Half Time Eligibility Error	7-44
7-38	Estimated Sampling Errors by Selected Categories: Number of Recipients with Insufficient Program Length Eligibility Error	7-45
7-39	Estimated Sampling Errors by Selected Categories: Number of Recipients with Nondegree Program Eligibility Error	7-46
7-40	Estimated Sampling Errors by Selected Categories: Number of Recipients with No Statement of Educational Purpose Eligibility Error	7-47
7-41	Estimated Sampling Errors by Selected Categories: Number of Recipients with No Financial Aid Transcript Eligibility Error	7-48
7-42	Estimated Sampling Errors by Selected Categories: Number of Recipients with Loan Default Eligibility Error	7-49
7-43	Estimated Sampling Error by Selected Categories: Number of Recipients with Invalid SAR Eligibility Error	7-50
7-44	Estimated Sampling Errors by Selected Categories: Number of Recipients with Not Parent School Eligibility Error	7-51

1
INTRODUCTION

This volume is one of a series that documents Stage Three of a project on Quality in the Pell Grant Delivery System. Other volumes concern the substantive findings of the project and recommendations for actions to correct the problems found.

In September, 1980, the Office of Student Financial Assistance (OSFA) of the U.S. Department of Education contracted with Advanced Technology, Inc., of McLean and Reston, Virginia, to conduct a three-year study to assess the accuracy and reliability of the Basic Educational Opportunity Grant (BEOG) Program and recommend administrative changes to improve it. Westat, Inc., of Rockville, Maryland, has served as a subcontractor to Advanced Technology throughout the study. In 1981, the name of the program was changed to the Pell Grant Program.

Pell Grants are entitlements available to students to support part of the cost of pursuing postsecondary education. The students, the schools they attend, and the programs in which they enroll must satisfy categorical eligibility requirements. Once these requirements have been met, the amount of the student's grant depends on the student's "need," which is a function of both the cost of the program and of the student's ability to pay that cost from student and family assets and income. The calculation of costs and the student/family contribution are prescribed by program formulae. Awards can be made in error if the student or the program is deemed to satisfy categorical criteria which they do not, in fact, meet, or if the student's need is miscalculated, or if the institution miscalculates the student's costs, enrollment status, or award. The money for awards is usually allocated from the U.S. Treasury to the institutions, which pay it to individual students or credit it to their accounts.

During Stage One of the study--the first year of the contract (1980-81)--Advanced Technology and Westat examined a national sample of Pell Grant recipients to determine eligibility and award calculation error. Westat drew the national sample

of 4,500 recipients and interviewed them and their parents about their eligibility and financial situation, examining documentation at the same time. Advanced Technology hired staff to visit the institutions attended by these students, examine the records on the students and any supporting documents on file, and interview the financial aid administrators. This data collection procedure was also followed in Stage Three--the third year of the contract (1982-83)--and is described below.

The Stage One study determined program-wide rates of discrepancy between actual grant awards and the awards that should have been made according to program rules and the documents examined, and attributed these discrepancies to institutions, recipients or their parents, and application processors. On the basis of these discrepancy rates, the study also identified error-prone groups of recipients. Finally, Stage One suggested feasible corrective management activities to reduce error rates for every area in which error rates were excessive.

During Stage Two (October, 1981, to December, 1982), Advanced Technology began the design of a quality control system for the Pell Program and made some error analyses and corrective action recommendations for specific features of related student aid programs. The Department of Education began installing corrective measures, including a requirement for greatly increased validation of Pell Grant applicants on a limited number of application items, rather than a small sample on more items. In 1983, the quality control system design component became a separate project.

Stage Three of the study (1982-83) has essentially been a replication of Stage One, with the objective of determining changes in program error over time, especially changes potentially brought about by the extended validation requirement. As in Stage One, Westat developed a sampling procedure and interviewed the sampled parents and students, and Advanced Technology visited institutions to examine documents in student files and interview financial aid administrators. During Stage Three Advanced Technology made a preliminary visit to each of 317 institutions to draw the sample of 4,109 students on site (during Stage One Westat had drawn the sample at its home office from lists supplied by the institutions) and to gather data for an assessment of compliance with the new validation requirement.

A findings report (Volume 1) and corrective action recommendations based on these data (Volume 2) are again accompanying this report on methodology (Volume 3). The contents of this volume consist of the procedures and methods of the study including sample selection, institutional visits, student and parent data collection, data processing, and data analysis. In addition, sampling response rates and detailed tables of the variance estimates for Volume 1 are included in a final chapter. Methodology that is deemed to be essential for understanding the findings of the report (Volume 1) is included in that volume. Thus, Volume 1 includes error definitions, strength of documentation, weighting and nonresponse adjustment, and experimental bias assessment in addition to findings on error in the Pell Grant program.

2
SAMPLE SELECTION

This chapter outlines the procedures used to select the sample of institutions for inclusion in the study and to select students at those institutions for interviewing. Two sampling techniques were used to select the students to be interviewed: one for students attending institutions distributing Pell Grants under the Regular Disbursement System (RDS), and one for students attending institutions using the Alternate Disbursement System (ADS).* The result was a self-weighting probability sample of all Pell recipients.

2.1 OVERVIEW OF SAMPLING FOR THE STUDY

2.1.1 The RDS Sample

The sample of students attending the RDS institutions was selected using a two-stage sampling procedure. The institutions were arranged in geographic clusters during the first stage of selection, with stratification by type of institution (less than two years, two to three years, and four years or more) and type of control (public, private, and proprietary) done at the second stage. The sampling frame for the RDS institutions was constructed from the Pell Grant Program Institution Universe File and the October 1981 Progress Report Error/Unreasonable File. Only active institutions within the contiguous 48 states were included.

Westat selected institutions according to the following steps:

- They ordered the file by geographic code, with clusters formed consisting of consecutive schools. Each cluster contained about eight schools.

---

*Additional detail on the sample design can be found in "Technical Report on Stage Three Sampling Procedures for Validation Evaluation Procedures," December, 1982.

- Clusters were sampled with probability proportional to total recipients in the cluster.
- They assigned an adjusted measure of size (AMOS) to each school within the sampled clusters, where:

  $AMOS_{ij} = IMOS_i/CMOS_j$

  i = school

  j = cluster

  IMOS = the institution measure of size

  CMOS = the cluster measure of size
- They ordered schools in sample clusters, ignoring the cluster identifier, by the following type-control strata:
  - School control (public, private, proprietary)
  - School type (offering 1 to 2 year programs, offering 2 to 3 year programs, offering 4 or more year programs)
- They sampled schools systematically from the ordered file with probability proportional to the adjusted measure of size (AMOS).

The student sample was drawn on site at each institution by data collectors hired and trained by Advanced Technology. At each institution, the data collector obtained a "list" of Pell Grant recipients from the institution and numbered them from 1 to n. Westat provided for each institution a list of the sequence numbers to be sampled, based on a fixed interval with random start. The intervals were calculated to give each student an equal probability of selection, and were based on the number of expected recipients. Sometimes the "list" consisted of Student Aid Reports (SAR) in a pile or file folder, the financial aid files of all the students at the institution, or other unusual record formats.

Data collectors occasionally found contaminated lists which included either recipients of other types of aid or people who had not, in fact, registered and received a Pell award at the institution. Where it was practical to do so, the data collector cleaned the list before selecting the sample. Where this was not practical, the list was used in its original form, resulting in oversampling. Checking files or records of the

sample enabled the data collector to eliminate from this oversized sample ineligibles who were not part of the study population. Any ineligibles still remaining in the sample were screened out prior to the student interview or the student record abstract.

2.1.2 The ADS Sample

Students attending institutions using the Alternate Disbursment System were selected from a back-up tape used by the Treasury to mail checks to Pell Grant recipients at ADS institutions. The ADS sample was not clustered by institution for two reasons. First, there was no easily available indicator of what school these students attend. Second, there are very few students in the ADS population compared to the RDS students; therefore, we expected a proportional sample of only about 20 students, minimizing any gains that might have been acquired in the conduct of the survey through clustering. The ADS student sample was selected according to the following steps:

- The Treasury tape was purged of duplicate records.
- The number of unique records in the unduplicated file was determined.
- A simple random sample of students was selected from the unduplicated file.

The ADS sample was selected using the same selection probabilities as the RDS sample, resulting in a self-weighting probability sample combining both categories of students.

2.1.3 The Control Group Samples

To provide a means of determining whether experimental bias existed, we selected two control groups, one from the sampled institutions (Institution Control Group or ICG) and one from both sampled and non-sampled institutions using the Computed Applicant Record (from the central processor) Control Group (CARCG). The ICG was selected from the sample lists used to draw the RDS sample. A fixed interval five times greater than the main sample interval was used with a new random start, resulting in a sample of about 800 students. Thus, institutions with a small

number of sampled students frequently had no control sample students. For 12 institutions, no list or reasonable facsimile was available and no control sample could be drawn.

The CARCG was drawn from the central processor's applicant file of more than seven million records. A sample of 20,000 applicant records was selected using a fixed interval and a random start. This sample was matched with the Pell Recipient History File to exclude non-recipients and separate students attending sampled and non-sampled institutions. The resulting file contained 9,051 recipients at non-sampled institutions and 1,997 recipients at sampled institutions.

2.2 DETAILED SAMPLING PLAN

The primary objective of the sample design was to choose a probability sample of students enrolled at educational institutions participating in the 1982-83 Pell Grant program. The statistically representative sample was designed to minimize the variability of characteristics among possible samples that can result from sample selection. The method used to achieve this goal was sample selection with probability proportional to a measure of size (MOS) that is presumed to be correlated with the statistics being estimated. In the institution sample selection, the measure of size used was the number of Pell Grant recipients reported by the institution in the October 1981 Progress Report. The October 1981 Progress Report was used for two reasons: (1) the October report satisfied the need for a sample of students receiving grants in the Fall; and (2) current year (1982-83) Progress Reports were incomplete.

2.2.1 Sample Size and Precision

The choice of sample size was determined, in large part, by the budget and time contraints of the study and the level of precision sought for sample estimates. The maximum number of institutions that reasonably could be contacted during the period provided for institutional data collection was about 300. In terms of the recipient sample, the budget permitted 3,600 completed interviews.

Based on the data available from the 1980-81 study (Stage One), basic assumptions on parameters that were not available at that time, we estimated the precision that can be expected with 300 institutions and an average of 12 students per

institution. A student characteristic such as absolute dollar error ($\bar{x}$) can be looked at as the product of two variables to be estimated from the survey: the average absolute dollar value of the error given that an error was made ($\bar{e}$); and the proportion of cases in error (P); thus

$$\bar{x} = P\bar{e}$$

The relvariance (square of the coefficient of variation) from a cluster sample may be approximated by:

$$V^2_{\bar{x}} = \frac{1}{m\bar{n}} \quad V^2 \quad 1 + \rho\ (\bar{n} - 1)$$

where:

$$V^2 = \frac{V^2_{\bar{e}} + (1-P)}{P}$$

$V^2_{\bar{e}}$ = relvariance, between students, of the amount of error;
P = proportion of cases in error;
m = number of clusters;
ρ = intra-cluster correlation coefficient; and
$\bar{n}$ = average cluster size.

Estimates of the coefficient of variation for several values of P are presented in Table 2-1. The following parameter values were assumed.

ρ = .5 (variance estimates from Stage One suggest that the intra-class correlation between students in a school is quite high)

$V^2_{\bar{e}}$ = 1 (derived from Stage One variance estimates)

$m\bar{n}$ = 3600

TABLE 2-1

EXPECTED COEFFICIENT OF VARIATION FOR AVERAGE ABSOLUTE DOLLAR VALUE OF ERROR, ASSUMING DIFFERENT VALUES OF P

	Proportion Of Students Who Had Some Error			
	P = .8	P = .6	P = .4	P = .2
Coefficient of variation	.052	.065	.085	.127

2.2.2 Design Objectives

In terms of survey implementation, the proposed sample design had the following objectives:

- Limit the amount of field travel
- Control the number of students and separate institutions selected
- Insure the representation of a variety of institutions

The procedures described below yielded a sample meeting these objectives.

2.3 SAMPLING FRAME: INSTITUTIONS USING REGULAR DISBURSEMENT SYSTEM

The sampling frame for the RDS institution sample was constructed from the Pell Grant Program Institution Universe File and the October 1981 Progress Report Error/Unreasonable File. The following steps were taken in the construction of the frame.

2.3.1 Selection of a Single Record Per Institution

A single record per institution was extracted from the Progress Report File, which contained progress reports for October and ad hoc or update reports. The initial October report was used unless an udpate was available. Branches of institutions reporting through a central office had no Progress Report record because the central

office accounted for their Pell Grant recipients. Details on the treatment of multi-branch campuses where the central office reports for all branches are given in a later section.

2.3.2 Selection of Eligible Institutions

After a single Progress Report record was extracted for each institution, these records were merged with the Universe File, by institution ID, and only in-scope institutions were retained in the frame. For the purpose of this survey, institutions are considered in-scope if they met the following conditions:

- They were in the coterminous United States (excludes Puerto Rico, Alaska, Hawaii and the Virgin Islands).
- They had RDS participation codes--position 240 on Universe File--of either 1 (participating, independent campus), 2 (central office for participating branch campus sytem), or 3 (branch campus participating through a central office).
- They had an eligibility code--position 239 on Universe File--of 1 (eligible for Pell Grant).
- They had an institutional status code--position 494 on Universe File--of 1 (active institution which may or may not be funded).

2.3.3. Imputation of Recipients (Measure of Size)

Eligible institutions which were on the Universe File but not on the Progress Report File--other than those reporting through a central office--were flagged for imputation of the number of recipients. Two hundred and forty-two institutions were in this group. Based on 4,676 institutions that had data on the number of recipients, undergraduate enrollment, institution type and institution control, we estimated a regression function to predict recipients as a function of the other three variables. In the estimation, we generated dummy variables for institution type and control categories and used the most general model, which included two-way and three-way interactions. The R^2 obtained was .72 (proportion of the variance in the dependent variable explained by the regression); a very good fit for cross-section data. Sixty-three institutions had no reported recipients and no undergraduate enrollment figure and thus imputation with the regression function was not possible. These institutions were listed for inspection, and since they were fairly small schools (such as

"Pedigree School of Dog Grooming"), they were assigned a value of 8 recipients to be used as measure of size.

2.3.4 Treatment of Campuses where the Central Office Reports for all Branches

Institutions where the central office reported for all the branches (Participation Code = 2) were handled as follows:

- For campuses broken down by branches in the "Education Directory - Colleges and Universities 1981-82", we allocated the recipients reported by the central office among the main campus and branches according to their respective enrollments; that is:

$$\text{Recipients}_B = (\text{Enrollment}_B/\text{Total Enrollment}) \times \text{Total Recipients}$$

- For other institutions--if the central office was selected, we called the school and determined if records of Pell Grant recipients were kept at the central office or at the individual branches. If records were kept at the central office, we sampled students from the central office files, and individual branches were not distinguished.

- For institutions that kept records at the individual branches, we obtained the breakdown of number of participants by branch and selected one branch within the institution, probability proportional to size (PPS), using the recipients at the branch as the measure of size.

2.3.5 Determination of Certainty Institutions

To obtain a final sample of about 300 cooperating schools--allowing for 2 percent out-of-scope institutions and 95 percent institution cooperation--requires an initial sample of 322 institutions. Institutions with reported recipients greater than the overall institution selection interval (total measure of size/322) were drawn into the sample with certainty.

The certainty cutoff was set at 3,740 recipients or approximately 70 percent of the school selection interval. That is,

$$\frac{\text{Total Recipients}}{\text{School Sample}} \times .70 = \frac{1{,}746{,}131}{322} \times .70 \doteq 3740$$

This certainty cutoff yielded 34 certainty institutions, 4 of which were central offices.

2.3.6 Selection of Noncertainty Institutions

The sample design for the noncertainty portion of the sample was basically a double sampling procedure with PPS selection of clusters and implicitly stratified PPS selection of all schools included in the sample clusters. The sample design called for:

- Ordering the file by geographic code and forming clusters of consecutive schools, from the ordered frame of about 8 schools each
- Sampling clusters with probability proportional to total recipients in the clusters
- Assigning an adjusted measure of size (AMOS) to schools within sampled clusters where:

  $AMOS_{ij} = IMOS_i/CMOS_j$

  i = school;

  j = cluster;

  IMOS = the institution measure of size; and

  CMOS = the cluster measure of size

- Ordering the schools sample clusters, ignoring the cluster identifier, by the following type and control strata:

  School control

  Public
  Private
  Proprietary

  School type

  Offering 1 to 2 year programs
  Offering 2 to 3 year programs
  Offering 4 year (or more) programs

- Sampling schools systematically from the ordered file with probability proportional to the adjusted measure of size (AMOS)

To limit the amount of field travel, we clustered institutions by geographically contiguous three-digit zip codes. (A previous step verified zip code and flagged those

requiring correction.) We defined a cluster as a set of 8 consecutive schools with total minimum measure of size to insure an average of 12 recipients per institution. The minimum measure of size was achieved by all clusters. The clustering process resulted in 632 clusters of 8 schools each, with an average of 2,470 recipients per cluster.

Sample of Clusters. Out of the 632 clusters in the frame, we sampled 72 clusters, PPS, with total number of recipients in the cluster as the measure of size (CMOS). The 72 sampled clusters contained 576 schools. None of the clusters sizes exceeded the sampling interval of 21,713 which would have required their selection with certainty at this stage.

Sample of Institutions. To insure adequate representation of different institutions, they were stratified by control (public, private and proprietary) and type (less than 2 years, 2 to 3 years, and 4 years or more). The measure of size for PPS selection was the ratio of the institution's MOS to the measure of size of the cluster from which the institution came (CMOS). That is,

$$AMOS_{ij} = IMOS_{ij}/CMOS_j$$

as defined above. The 576 institutions in the 72 sample clusters were sorted by the type and control strata defined above and a systematic sample of 288 institutions was selected with probability proportional to the adjusted measure of size (AMOS). A considerable number of institutions (171 or 59.3%) had a measure of size larger than the sampling interval and were conditional certainty institutions; these were removed from the frame before the 117 noncertainty institutions were drawn.

The resulting distribution of the sample among type-control strata was proportional to the distribution of the universe measure of size among the strata. This is illustrated in Table 2-2.

2.4 SELECTION OF STUDENTS FROM SAMPLE RDS INSTITUTIONS

To obtain a sample of about 3,600 Pell Grant recipients--300 institutions with an average of 12 students per school--required an initial sample of about 4,040 recipients, assuming a 90 percent combined institution/recipient response rate. The overall

TABLE 2-2

DISTRIBUTION OF RECIPIENTS IN INSTITUTION UNIVERSE
COMPARED TO SAMPLED INSTITUTIONS

Institution Type	Public		Private		Proprietary		Total	
	Value	Percent	Value	Percent	Value	Percent	Value	Percent
Less than 2 years								
Universe MOS	10,343	.6	3,126	.18	92,704	5.3	106,173	6.1
Sample count	2	.6	--	--	22	6.8	24	7.5
2 to 3 years								
Universe MOS	451,443	25.9	34,663	1.9	46,416	2.7	532,522	30.5
Sample count	88	27.3	10	3.1	14	4.3	112	34.7
4 years or more								
Universe MOS	759,500	43.5	342,776	19.6	5,160	.3	1,107,436	63.4
Sample count	111	34.5	74	23.0	1	.3	186	57.8
Total								
Universe MOS	1,221,286	70.0	380,565	21.7	144,280	8.3	1,746,131	100.0
Sample count	201	62.4	84	26.1	37	11.4	322	100.0

2-11

recipient sampling fraction required to achieve an equal probability sample of the desired size is given by:

$$f = \frac{n_I}{TMOS} = \frac{4040}{1746131} = \frac{1}{432.21}$$

where

f	=	overall sampling fraction
TMOS	=	total number of recipients (total measure of size) computed from the October 1981 Progress Report File
n_I	=	the required initial sample of recipients

Consequently, the weight associated with each sample recipient--the reciprocal of the overall sampling rate--was 432.21.

Within any institution, the probability of selecting a given student was equal to the overall sampling fraction. Thus, the products of the probabilities of selecting a particular cluster, institution within the cluster, and recipient within the institution were equal to 1/432.21. The sample, then, was effectively self-weighting, with no adjustments necessary for institutional characteristics.

2.4.1 Sampling Frame

The sampling frame was the list of Pell Grant recipients, at each of the sample institutions, from which the recipient sample was drawn. At the time we contacted the school to set up an appointment for the data collector's visit, we requested that the school prepare a list of current and pending recipients to be used as a sampling frame.

Recipients were selected from the sampling frame systematically with equal probability. For this purpose a sampling worksheet was produced for each institution. The sampling worksheet included all the information necessary for drawing the sample, such as sample interval and pregenerated selection numbers.

After the data collectors obtained a list that included all Pell Grant recipients eligible for the study, they numbered the students on the list sequentially. Then, they

selected those students with sequence numbers corresponding to the selection numbers given on the sampling worksheet.

2.4.2 Weighted Estimates

The sample of Pell Grant recipients from RDS schools is an equal probability self-weighting sample; before adjustments for nonresponse all students have the same weight. The procedures used for determining nonresponse weights have been described in Volume 1. The sample selection procedure was designed to be used in providing estimates of student characteristics. Thus, estimates for institutions are not recommended.

3
INSTITUTIONAL VISITS

Institutional visits were conducted in both the fall of 1982 and the spring of 1983. The fall visits had two purposes: to draw the sample of students who would, as well as their parents, be interviewed in the spring; and to compile minimal information on newly-instituted institutional validation activities. The results of the latter investigation were delivered to the Department of Education in December, 1982. The methods for training, scheduling, and managing the fall data collectors were generally similar to those reported here for the spring visits.

Two previous reports on data collection have been submitted to the Department of Education, one on training and qualifications of data collectors and one on the data collection itself.* Those reports contain copies of forms used, interview scripts, and resumes of the data collectors. This chapter is a compilation of all relevant procedural and methodological sections from those reports.

3.1 DATA COLLECTORS

The data collectors were the key to a successful spring data collection. Given the complexity of the Pell Grant Program and the variety of documents and institutional recordkeeping systems likely to be encountered, financial aid experience was essential. Since the data collectors would be working alone all over the country, detailed supervision of them would be impossible; therefore, they had to be reliable and mature enough to be able to make their own decisions in many situations.

*Advanced Technology, Inc., "Report on Training and Institutional Data Collectors for Stage Three of the Pell Grant Quality Control Study," April, 1983, and "Report on Spring, 1983, Data Collection for Stage Three of the Pell Grant Quality Control Study," June, 1983.

3.1.1 Recruitment

We began recruiting data collectors with the people who had performed successfully in our two previous data collections in the spring of 1981 and the fall of 1982. None of the 1981 data collectors was available for data collection in 1983. However, three of the fall, 1982, field staff were available and were rehired.

We advertised our field staff needs in four newspapers and periodicals: the Newsletter of the National Association of Student Financial Aid Administrators (NASFAA); the Chronicle of Higher Education; The New York Times; and The Washington Post. The NASFAA Newsletter provided the greatest number of applicants with the type of experience we were seeking (at least 12). The two newspapers advertisements generated few or no useful responses. When one of the people who had been offered a position declined at the last moment, a well-qualified replacement was hired through the personal contacts of Deputy Project Director Richard Tombaugh, former executive secretary of NASFAA and former director of the Student Financial Assistance Training Program.

3.1.2 Qualifications

The spring, 1983 data collectors were the best qualified of the three groups used so far. Nine of the 15 people initially sent into the field had been financial aid directors, 2 at large state institutions; 3 others had over 10 years experience each in financial aid administration. The two data collectors without financial aid experience had performed successfully as data collectors in the fall of 1982. The 15 original field staff included 4 with bachelor's degrees, 10 with master's or first professional degrees, and 1 doctorate; 1 of those with a master's was an active doctoral candidate. The degrees in most cases were in education.

Therefore, the data collectors were well-equipped by training and experience to deal with the variety of documents and record systems they would encounter in the field, to explain interview questions to financial aid directors, and to understand answers couched in technical financial aid terminology. They also could find relevant data in student files quickly and accurately. The training then concentrated on ensuring consistent use of the data collection instruments, with little need for instruction on the basics of the Pell Grant Program.

3.2 SCHEDULING

The principal objective of the scheduling was to minimize travel costs while visiting all the sample institutions within the six-week data collection period with only 15 interviewers. We drew up an ideal master schedule to achieve this goal and then adjusted it as necessary when we could not visit on our preferred dates.

3.2.1 Master Schedule

The first step in drawing up an ideal master schedule was to estimate how much time was needed for each institutional visit. Since the interview had a fixed cost for each visit, the time allowed depended on the number of students sampled. We estimated that each interview would take an hour and that each student abstract would require about 20 minutes initially, but slightly less for larger numbers. Therefore, we estimated that each data collector could handle 14 students during the first and second days at a school, 16 the third day, and 20 on the fourth day. We broke these estimates down by half days as follows:

Number of Sampled Students	Number of Days
1 - 7	½
8 - 14	1
15 - 21	1½
22 - 28	2
29 - 36	2½
37 - 44	3
45 - 54	3½

Once we had estimated the time required for each institutional visit, we grouped institutions into clusters that would each require one data collector one week to visit. We had to allow time to travel between cities in making these clusters. In some cases, a data collector could stay in a city for a whole week, even being scheduled to do two half-day schools in one day. Although the institutional sample had been clustered for efficiency in both student and parent interviewing and institutional

visits, certainty schools--those whose large number of Pell recipients made their selection for the sample certain--could be anywhere and some "clusters" were scattered geographically. Therefore, some of the institutions we had to schedule were in fairly inaccessible or remote places, like Susanville, California; Trinidad, Colorado; and Spearfish, South Dakota; others were located hundreds of miles from the next nearest institution's location, as was Tuscon, Arizona. We usually allowed a full day to get from or to one of these places. We assumed that a weekend was sufficient to get from any one-week cluster to another. Occasionally, the schools were arranged geographically in such a way that we could not avoid giving an interviewer an open date even though a whole day was not needed for travel.

After clustering the institutions into one-week work loads, we grouped the clusters into 15 regions of 6 clusters each. Each region represented one data collector's assignment. These regions are shown in Figure 3-1. A major consideration in this grouping was the creation of loops which would enable a data collector to use the same rental car for six weeks and return it to the place of origin, saving drop charges, or taking advantage of round-trip air fares.

Once institutions had been grouped into weeks and weeks into regions, we began assigning preferred dates to the institutions. We tried to designate two alternate dates for each school, but this was not always practical. Where we had several schools to visit in a single city or within a few miles of each other, we could rotate the preferred dates through several days or a week easily. In other areas, a week's work could be treated as a loop itself, with a preferred schedule specifying travel in one direction and an alternate schedule involving travel in the reverse direction. The final master schedule consisted of a list of all 317 participating institutions with one preferred and one or two alternate dates for each.

3.2.2 Schedule Confirmation

We used Westat's telephone facilities and staff to call all the institutions to secure agreement on the date of our visit. Westat's telephone center includes sophisticated switching equipment which selects the cheapest telecommunications system (MCI, SPRINT, AT&T, etc.) for each call, sound-proofed carrels for interviewers, and monitoring equipment. Westat has available a corps of telephone interviewers who are experienced in following a written script, recording answers, and tracking individual questionnaires to completion.

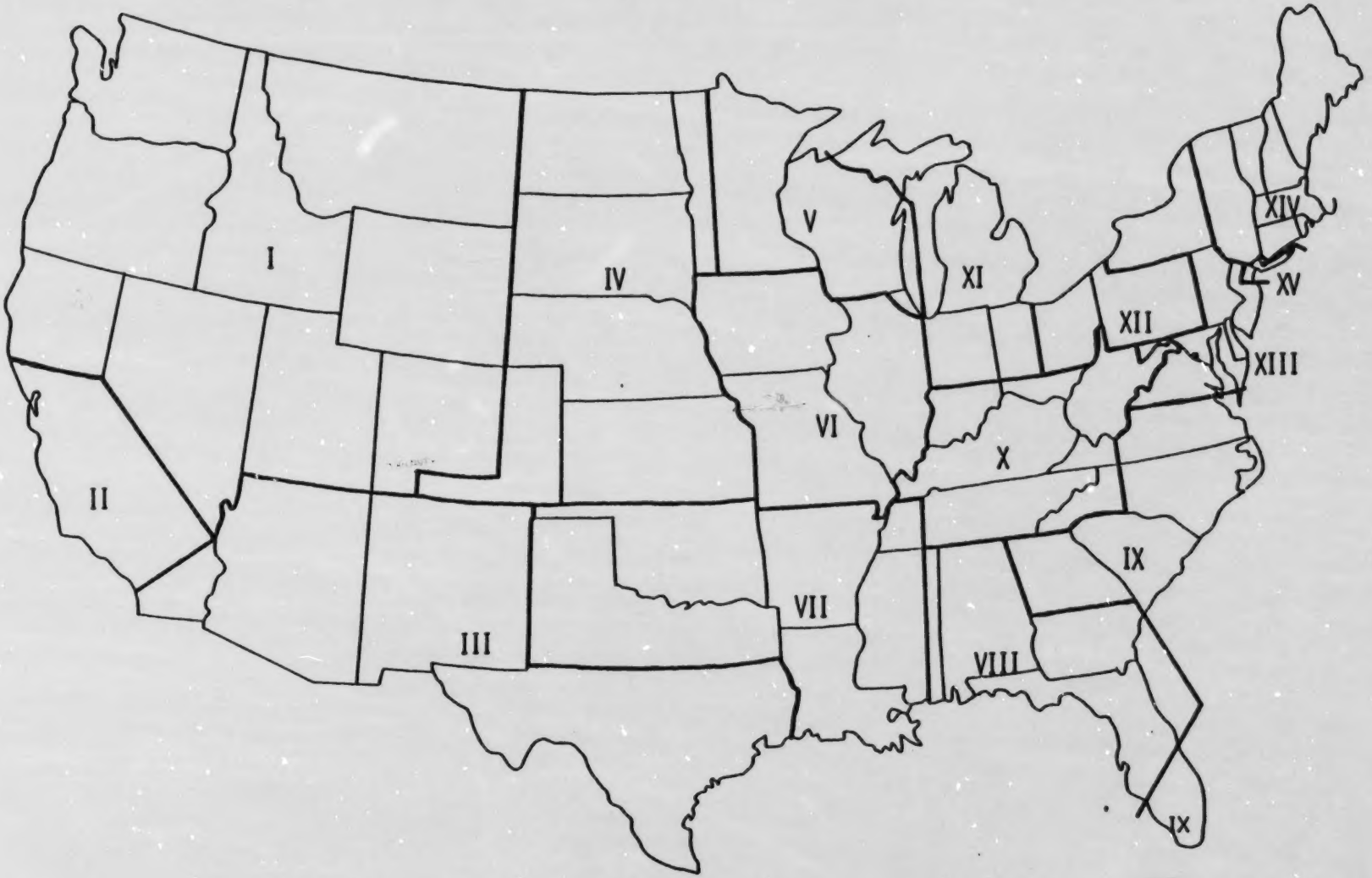


FIGURE 3-1

DATA COLLECTION REGIONS

To confirm each scheduled visit and collect some other information for our data collectors, we developed a script for the Westat telephone interviewers (Figure 3-2). If an institution would not accept a visit on the preferred date or any alternate date, the interviewer asked what dates would be acceptable. The script also gave us an opportunity to confirm that the data we required would be available, to reiterate our requirements, and to give the financial aid director at each school advance notice of some questions for which he might have to compile statistics.

Westat telephone staff immediately reported to our data collection manager (who had a temporary office at Westat's telephone center during the scheduling period) every case in which a school would not agree to a visit on the preferred date. These situations required moving other schools to alternate dates or wholesale rescheduling. We also had a script for rescheduling calls. Whenever possible, rescheduling was limited to schools which had not yet been reached, to minimize the number of changes and the number of times individual directors were called back.

Westat trained the scheduling interviewers. An Advanced Technology professional staff member was assigned to monitor the scheduling calls on a rotating basis to cover all the interviewers; when he had time, our data collection manager also monitored the scheduling calls. The monitors brought minor problems to the attention of the interviewers immediately; they encountered no major problems.

A few schools which were difficult to reach were called directly by Advanced Technology staff after the initial scheduling period had formally closed. After all schools had agreed to visit dates, we compiled an individual schedule for each data collector listing the dates, places, and times of the visits, the names of the financial aid directors, and the numbers of students in the main and control group samples.

3.3 TRAINING

Since all the field staff were experienced financial aid administrators or had been data collectors for the fall, 1982, validation evaluation, little training was needed on the Pell Grant Program. Therefore, it was possible to devote an entire week to interviewing techniques and to filling out correctly a Student Record Abstract (SRA) for each student in the main sample. Figure 3-3 is the training schedule.

INSTITUTION CODE _____

INSTITUTION NAME _____

_____

City State

TELEPHONE NO. ( ) _____

PELL GRANT SCHEDULE CONFIRMATION
SPRING, 1983 INSTITUTIONAL VISITS

1. Good (morning/afternoon), (Mrs./Mr.) _____. I am _____ _____, from Westat, Inc., in Rockville, Maryland. Last fall a representative of Advanced Technology, Inc., visited your institution to collect information on a sample of your Pell Grant recipients as part of the Pell Grant Quality Control Study. A few weeks ago Dr. Ernst Becker, head of the Division of Quality Assurance in the Office of Student Financial Aid, sent you a letter describing the spring data collection and the activities Advanced Technology will be carrying out at the sample institutions. Have you received that letter and had a chance to read it?

   _____ YES (GO TO 3)

   _____ NO, DON'T KNOW, CAN'T RECALL (CONTINUE BELOW)

2. During April and May representatives of Advanced Technology, Inc., the prime contractor for the study, will be visiting each of the sample institutions to interview their financial aid directors in more detail on validation and ways to reduce payment error in student financial aid programs. They will again be abstracting information from the files of the students selected last fall. Those students and their parents are also being interviewed individually; many have already been contacted by field staff.

3. For planning purposes, we have established a tentative schedule to visit all the institutions this spring and hope that most institutions will cooperate so we can meet the Department of Education's deadline for our report. I have a checklist of items to ask you regarding our visit to your institution.

FIGURE 3-2

SITE VISIT SCHEDULING SCRIPT

3a. Our interviewer is presently scheduled to visit your institution on _____ at _____. Is that date and time acceptable to you?

_____ YES (GO TO 3c)

_____ NO (IF DATE UNACCEPTABLE, GO TO 3b; IF TIME ONLY UNACCEPTABLE, CONTINUE BELOW)

What time can our visitor arrive to get in a full day's work?

(RECORD TIME) _____ (GO TO 3c)

3b. Because our interviewers have so much travelling to do, it is important that we be able to visit all the institutions in one city or area on a single visit. Pending confirmation from the other institutions in your area, _____ or _____ would be good alternate dates for us. Would they be acceptable to you? (IF "YES," GO TO 3c; IF "NO," CONTINUE BELOW.)

Although we would prefer to interview the financial aid director at each institution at the beginning of our visit, we could talk to an associate or assistant director who can answer authoritatively about your institution's policies and practices regarding Pell Grant validation and administration and packaging of student aid; or we could begin the visit with the student record reviews and talk to you later in the day. Could we use either of these alternatives to keep to one of the dates I have suggested?

_____ YES (RECORD WHICH DATE AND TIME) _____

_____

_____ NO (CONTINUE BELOW)

Keeping in mind that we have to keep a very tight schedule, what would be the closest acceptable date to the date I originally suggested? I originally suggested _____. (RECORD ALTERNATE DATE) _____.

FIGURE 3-2 (cont'd)

SITE VISIT SCHEDULING SCRIPT

3c. Are there any local holidays or school vacation periods that might interfere with the date we have agreed to?

____ YES (CONTINUE BELOW)

____ NO (GO TO 4d)

Is that a local holiday which might affect other institutions in your area, or is it specific to your institution?

____ HOLIDAY, MIGHT AFFECT OTHER INSTITUTIONS

____ SPECIFIC TO OWN INSTITUTION

Again, keeping in mind our need to visit several institutions when we are in you area what would be the best time for our visit that would be close to the date I originally suggested? That date was _____.

(RECORD DATE) _____

4. To determine enrollment status and the amounts and dates of Pell Grant disbursements, our interviewers often have to visit the registrar and bursar, as well as the student aid office. Are both the registrar and bursar located in the same building as your office?

____ YES (GO TO 4b)

____ NO (CONTINUE BELOW)

4a. Where is the registrar's office?

(RECORD ANSWER) _____

Where is the bursar's office?

(RECORD ANSWER) _____

We would appreciate your notifying the bursar and registrar of the visit, just in case our researcher needs to talk to them or check their records.

5. Are all the records for your institution's students on your campus, o are any of them kept elsewhere, at a central office or branch campus?

____ ALL RECORDS KEPT HERE (GO TO QUESTION 6)

____ SOME OR ALL RECORDS KEPT ELSEWHERE (CONTINUE WITH 5a BELOW)

FIGURE 3-2 (cont'd)

SITE VISIT SCHEDULING SCRIPT

5a. What records are kept elsewhere?
(RECORD ANSWER) _____

_____

5b. Where are those records kept? (RECORD NAME AND ADDRESS OF OFFICE, INSTITUTION, CONSULTANT, ETC.)

_____

_____

_____

_____

5c. Whom should we contact to get access to those records?
NAME _____

TELEPHONE NUMBER _____

5d. Are the records there complete, or would we have to visit your office or campus also?

_____ YES, RECORDS THERE ARE COMPLETE
_____ WOULD HAVE TO VISIT THIS OFFICE ALSO

6. Will enrollment and Pell disbursment records be open and available to our visitor on (DATE OF SCHEDULED VISIT) _____?

_____ YES (GO TO 7a)
_____ NO, PELL RECORDS NOT OPEN ON THAT DATE (GO TO 6a)
_____ NO, CANNOT PERMIT OUTSIDE INSPECTION (GO TO 6b)

6a. Again, keeping in mind our need to visit several institutions when we are in your part of the country, what would be the date(s) closest to the date I originally suggested on which we could visit. That date was _____.

(RECORD DATE) _____

FIGURE 3-2 (cont'd)

SITE VISIT SCHEDULING SCRIPT

6b. As the earlier correspondence explained, we are under contract to the U.S. Department of Education to complete this study, and, under existing regulations, information about a student's financial aid may be disclosed to authorized Education Department representatives without the consent of parents or students. We did examine Pell Grant records last fall for the students we selected then. Is there an official reason or new institutional policy which would require you to close the records?

YES (RECORD REASON)

_____

_____

_____

(CLOSE INTERVIEW AND REFER INSTITUTION TO PROJECT DIRECTOR.)

7a. Can you suggest where our researcher should park his or her car and note any special procedures to be aware of, such as checking in with campus security?

_____

_____

_____

7b. Is there any particular nearby hotel or motel which you recommend to people from out of town who are visiting your school? (DO NOT PROBE FOR ANSWER TO THIS QUESTION) _____

_____

_____

FIGURE 3-2 (cont'd)

SITE VISIT SCHEDULING SCRIPT

8. Before ending the call, let me just give you some advance notice concerning a couple of items our interviewer will be asking about which you might have to look up. We will need to know how many students were selected by you and by the U.S. Department of Education for Pell Grant validation, how many you asked for citizenship or residency documentation, and how many corrections were made on the basis of Veterans Administration information. We will also be asking about the effects of validation in terms of the number of students whose payments were delayed and how long the average delay was.

(IF RESPONDENT ASKS FOR DETAILS OR DEFINITIONS, READ THE FOLLOWING:)

Our researcher, who will be an experienced financial aid administrator, will have the details. We just want to let you know in a general way about a couple of questions that might require you to look up some numbers in order to answer correctly.

9. Thank you for your help. Let me review the date we have agreed to: Our researcher will arrive on _____ at _____. Is this your understanding?

_____ YES

_____ NO (RESOLVE)

_____

_____

_____

Our researcher will call you from the field a few days before arrival in order to confirm the visit.

FIGURE 3-2 (cont'd)

SITE VISIT SCHEDULING SCRIPT

TRAINING PROGRAM

PELL GRANT QUALITY CONTROL STUDY

Monday, April 4

AGENDA

8:15 - 8:30	Coffee
8:30 - 9:15	General Welcome, Introductions, and Overview of Training Week - Mr. Richard Tombaugh, Advanced Technology
9:15 - 9:30	Introduction to Advanced Technology - Dr. Alex Ratnofsky, Advanced Technology
9:30 - 9:45	Introduction and Background of Pell Grant Quality Control Study - Mr. Ernst Becker, Department of Education
9:45 - 10:00	Overview of Stage III Quality Control Activities - Ms. Carol Miller, Advanced Technology, and Mr. Bob Learmonth, Westat
10:00 - 10:15	Overview of Data Collectors' Tasks and Responsibilities - Richard Tombaugh
10:15 - 10:30	Break
10:30 - 12:15	Conducting the Institutional Interview With the FAA - Richard Tombaugh
12:15 - 1:00	Lunch
1:00 - 1:15	Review of Training Manual - Ms. Tracy O'Connor, Advanced Technology
1:15 - 3:00	Detailed Item-by-Item Examination of Institutional Interview Form - Mr. John Neely, Advanced Technology
3:00 - 3:30	Break
3:30 - 4:30	Demonstration of Institutional Interview - Carol Miller and Richard Tombaugh
4:30 - 5:30	Review Answers from Demonstration Interview; Group Practice: Beginning the Interview, Answering Respondent Questions, and Asking the Interview Questions - Carol Miller and Richard Tombaugh

FIGURE 3-3

TRAINING SCHEDULE

Tuesday, April 5

8:15 - 8:30	Coffee and ID Photos
8:30 - 9:30	Role Playing of Institutional Interview with Trainees in Pairs
9:30 - 10:30	Acting Out of Institutional Interview: Trainers Playing Difficult Respondents with Trainees as Interviewers to Practice Probing and Handling Difficult Situations -Carol Miller, Richard Tombaugh, and Dr. Albert Parker, Advanced Technology
10:30 - 10:45	Break and ID Photos
10:45 - 12:15	Detailed Item-by-Item Examination of Student Record Abstract and Specifics of Completing Abstracts - John Neely
12:15 - 1:00	Lunch
1:00 - 2:00	Distribution of Hypothetical Student Financial Aid Record File; Perusal of Sample Forms Likely to be Found in Student Aid Files; Review of Federal Tax Forms - Richard Tombaugh
2:00 - 3:00	Practice Completing a Student Record Abstract Using Data from Hypothetical Student Files
3:00 - 3:15	Break
3:15 - 4:15	Practice Completing a Second Student Record Abstract
4:15 - 5:00	Abstracting the Atypical Student Record - Richard Tombaugh
5:00 - 5:30	Completion of Business Forms and Associated Paperwork - Tracy O'Connor and Janice Grant

Evening Assignment: Student Record Abstract (Practice Three)

FIGURE 3-3 (cont'd)

TRAINING SCHEDULE

Wednesday, April 6

8:15 - 8:30	Coffee
8:30 - 9:15	Review of Answers for Tuesday Evening Assignment - Richard Tombaugh
9:15 - 10:15	Practice Completing a Fourth Student Record Abstract
10:15 - 10:30	Break
10:30 - 11:00	Discussion of Institution Control Group Activities - Albert Parker
11:00 - 12:00	Role Playing of Resolution/Exit Interview with FAA - Carol Miller and Richard Tombaugh
12:00 - 1:00	Lunch
1:00 - 2:00	Explanation of Shipping and Receiving of Data Collection Materials, Editing and Post-Interview Procedures - John Neely
2:00 - 2:30	Explanation of What Happens to Completed Data Collection Forms Upon Return to Project Office: Log-in of Forms, Interview Verification, Coding, Keypunching, and Data Processing - James Southwick, Charles Schueller, and Dr. Dan Geller, Advanced Technology
2:30 - 3:30	Explanation of Travel Arrangements - - Cash Advances, Hotel Reservations, Airline Tickets, Car Rental Procedures, and Expense Reports; Discussion of Field Reporting and Anticipated Problems - Tracy O'Connor and Albert Parker
3:30 - 3:45	Break
3:45 - 4:00	Description of Thursday, April 7, Field Practice - Richard Tombaugh
4:00 - 5:00	Practice Completing a Fifth Student Record Abstract

FIGURE 3-3 (cont'd)

TRAINING SCHEDULE

Thursday, April 7

9:00 - 12:00	Review of Pilot Data Collection for Campus-Based Programs - Richard Tombaugh (Justice, Alvarez, and Finch only)
All Day	Field Practice at the Following Area Institutions:
	American University Washington, D.C.
	Computer Learning Center (afternoon only) Springfield, VA
	George Washington University Washington, D.C.
	Georgetown University Washington, D.C.
	Montgomery Community College Rockville, MD
	Montgomery Community College Takoma Park, MD
	Howard University Washington, D.C.

FIGURE 3-3 (cont'd)

TRAINING SCHEDULE

Friday, April 8

8:15 - 8:30	Coffee
8:30 - 10:00	Detailed Review of Previous Day's Field Practice -- Discussions of Experience, Answers to Trainee Questions, Advice on How to Handle Problem Situations - Richard Tombaugh
10:00 - 10:15	Summary - Carol Miller
10:15 - 11:45	Confirmation Calls to First Week's Institutions
11:45 - 12:45	Lunch
12:45 - 2:00	Issuing of Cash Advances; Trip to Bank for Traveler's Checks
2:00 - 3:00	Recap of Training Week; Completion of Paperwork; Answering of Questions

FIGURE 3-3 (cont'd)

TRAINING SCHEDULE

3.3.1 Manuals

We produced two instruction books for the data collectors: a Data Collector Training Manual and Question-by-Question Specifications. We also distributed to the data collectors for reference in the field the Pell Grant Validation Handbook and two booklets published for OSFA's Student Financial Assistance Training Program: Aid Administrator's Guide to IRS Forms and Schedules and Index of Regulations. The Training Manual covered general data collection procedures, specific procedures for this study, and travel and accounting arrangements, but not how to code specific items on the data collection forms; that was the province of the Question-by-Question Specifications.

3.3.2 Interviewer Training

Since our data collectors had considerable experience in reviewing Student Aid Reports (SARs), income tax forms, and other documents in student aid files, but little or no experience in conducting structured interviews, we devoted a greater proportion of the training to the interview than the amount of time required for it in the field would suggest.

During the first day of the training week, we introduced the data collectors to general interviewing procedures with a sound and slide presentation developed by Westat for interviewer training, made a detailed, item-by-item presentation of the institutional interview form, and staged a demonstration interview with senior project staff playing the parts of interviewer and financial aid director. The second day of training included opportunities for each data collector to practice an interview, with another data collector playing the role of financial aid director, and a demonstration by project staff of how to handle three types of difficult situations: a hostile, uncooperative respondent, a voluble respondent who wandered off into long but irrelevant answers, and a taciturn respondent who had to be asked to expand or explain brief, cryptic answers. We also staged a demonstration of the exit interview in which data collectors were to attempt to resolve certain problems which they had uncovered when they filled out the Student Record Abstracts (SRAs). This demonstration was important because there was no specified form or script for this exit interview.

3.3.3 Student Record Abstract Training

Although our field staff had plenty of experience with the student aid forms and validation documents from which they were to collect the data on individual students, they needed training in filling out and editing our forms so that everyone would collect the same information in a format that could be coded easily at the project office. Therefore, we went over the SRA in detail, one item at a time, defining exactly what information we wanted, explaining what documentation we desired, discussing possible problems, and answering questions. The most important part of the training was a series of exercises in which we gave the data collectors student aid files for fictional students from which to practice filling out SRAs. We then reviewed the data and how they should have been entered on the form. Mr. Richard Tombaugh also conducted a special training session on some of the unusual situations which might be encountered in the field.

3.3.4 Site Visit Practice

On the fourth day of the training week, we sent the data collectors in groups of two or three to seven local institutions which were not included in the sample; each group was accompanied by a member of the project staff. The purpose of these visits was to give the data collectors an opportunity to practice the interview and completion of several SRAs under conditions which more closely approximated what they would encounter in the field.

We had arranged beforehand to interview the financial aid director or Pell Grant manager at each practice site. Each of the two or three data collectors did part of the interview. We had also asked the financial aid director to pull several files at random for each data collector. These were not a scientific sample but did provide practice with real files and, at some sites, computerized record systems.

On the day after the practice, the last day of the training, we held a discussion and evaluation of the practice at which the data collectors shared the lessons they had learned and the project staff answered questions.

3.4 FIELD SUPERVISION

Extensive field supervision of the data collection staff ensured that we became aware of problems and resolved them promptly. Field supervision included regularly scheduled calls to the field supervisor from the data collectors; unscheduled calls to the field supervisor for the resolution of specific problems; field visits to all data collectors by senior project staff; validation calls to a sample of the institutions visited by each data collector; and calls to data collectors for resolution of problems revealed by editing.

3.4.1 Scheduled Calls

Each data collector had a regular time at which to call the field supervisor each week, regardless of time zone or whatever else the data collector was doing. (Exceptions were made only for interviews which interfered with the call because of time zone differential.) The scheduled call was an opportunity for the data collector to raise any problems which had not required immediate consultation with the project staff, including prospective problems uncovered by calls to institutions to confirm appointments. It also gave the field supervisor a chance to discuss problems with the data collector, especially those which had appeared during editing at the project office but were not serious enough to demand an immediate call, such as items omitted on the SRA which required a call from the data collector to the institution. The field supervisor also informed the data collectors about errors in the completion of SRAs which did not require calling the institutions for further data, such as inadequate field editing. The coders had forms for this purpose which they could place in each data collector's file. The field supervisor had a checklist of potential problem areas to review with the data collectors each week.

3.4.2 Technical Liaison

Calls from the data collectors to the field supervisor were more frequent than regularly scheduled calls, especially during the early weeks of the study. These dealt with unique or unusual situations at particular institutions which had not been covered during training, and concerned the SRA almost exclusively. How to answer a particular question often depended on the purpose of the question or how the data would be used in analysis, so the field supervisor checked often with the manager for data analysis or other project analysts.

3.4.3 Monitoring

A member of the project management or analysis staff visited every one of the data collectors during the first eight days in the field to observe how they interviewed financial aid directors and completed SRAs. The visits were made by the project manager, the deputy project manager, the manager for analysis, and a project analyst. They monitored the new data collectors (those who had not worked on the project in the fall of 1982) during the first week, and the three returning data collectors on the first three days of the second week.

During these monitoring visits, the monitors answered questions which had not arisen during training, corrected some minor errors in completion of the SRAs, and made other suggestions to the data collectors. They found no systematic errors or any problems comparable to those which had led us to pull a data collector from the field during the fall.

3.4.4 Validation

Monitoring visits was too expensive to do more than once as long as no major problems were discovered. As a continuing check on the performance of the data collectors and a source of feedback from a different perspective, we instituted telephone validation. In the second week of the field period, a professional project staff member called every one of the financial aid directors who had been visited during the first week and administered a brief questionnaire which included an open-ended item about the data collector's performance (Figure 3-4). Each week thereafter, we called about half of the financial aid directors who had been visited in the previous week.

No problems were uncovered by this monitoring. The responses to the final open-ended question praised the professionalism, knowledgeability, and cooperativeness of the data collectors. None of the financial aid administrators who were interviewed raised any problems with the data collectors or objected to the way they did their jobs.

PELL GRANT QC PROJECT
SPRING, 1983 DATA COLLECTION

INTERVIEWER VALIDATION REPORT

Interviewer _____

Institution Visited _____

Financial Aid Administrator _____

Telephone Number _____

Date of Visit _____

Validation Calls (Enter call-back time in next column)

Date: ____ ____ ____ ____ ____ ____

Time: ____ ____ ____ ____ ____ ____

Hello, this is (YOUR NAME) from Advanced Technology, Inc. In order to assure the quality of the data collected for the Pell Grant study we are doing for the Department of Education, we are calling the participating institutions to evaluate the performance of our field personnel.

On _____ (Date of visit), _____ (Name of Interviewer) was scheduled to interview you and collect some data from some of your student files.

1. Did (he/she) arrive on time?

   1. Yes
   2. No

2. Did (he/she) present (his/her) credentials?

   1. Yes
   2. No

3. Did (he/she) conduct the interview with you in a professional manner?

   1. Yes
   2. No → What in particular did you find unprofessional about (his/her) conduct? _____

FIGURE 3-4

INTERVIEWER VALIDATION REPORT

4. Was (he/she) able to answer any questions you or your staff had about the study?

   1. Yes
   2. No ⟶ What was the question (he/she) could not answer? _____

   _____

5. Was (he/she) careful with your files and cooperative with your staff?

   1. Yes
   2. No ⟶ What problems did (he/she) cause? _____

   _____

6. Did (he/she) conduct an exit interview with you or one of your professional staff, or offer to?

   1. Yes
   2. No

7. Do you have any further comments about (his/her) performance, or any suggestions to pass on to (him/her)? (RECORD COMMENTS)

   _____

   _____

   _____

   _____

FIGURE 3-4 (cont'd)

INTERVIEWER VALIDATION REPORT

3.4.5 Problem Resolution

Whenever possible, data collection problems revealed by editing of SRAs were resolved through calls to the responsible data collectors. Sometimes they remembered the circumstances at a particular institution and were able to provide additional information immediately. In other cases, they had to call the institutions themselves. Having the data collector, rather than a member of the central project staff, call the institution had two advantages. First, the data collector was already known at the institution, knew which specific office or staff member to call, and knew the records system. Second, going back to the data collectors emphasized to them the seriousness of getting complete and accurate data and created an incentive for avoiding future mistakes.

The most common problem which required recontacting the school was ambiguity in cost-of-attendance data. Some schools had unusual ways of calculating this figure which were not absolutely clear to data collectors during the visit. In a few cases, necessary SRA data had not been reported in full.

3.4.6 Financial Reporting and Cost Management

Advanced Technology instituted a set of strict cost-reimbursement policies and explained them to the data collectors during training. Data collectors were required to submit comprehensive weekly expense statements with receipts and daily expense logs. When these reports were late, pay checks were not forwarded to data collectors' accounts. In a few instances, the field supervisor asked about unusual expenses during weekly calls; he also notified data collectors about non-reimbursable expenses at that time. Non-reimbursable expenses were rare and consisted mostly of personal telephone calls which had gone over the daily limit.

We required data collectors to notify us of travel changes beforehand. Most changes were to the financial advantage of the project. Because we had recruited data collectors nationally, we had several who were working near their homes and found it convenient to go home on weekends, saving us hotel bills and reimbursement for meals. Once they got to the cities they would be visiting, some data collectors found hotels that were both closer to their schools and cheaper than the ones our

travel agent had been able to reserve at a distance. In general, the field staff were quite cost-conscious and were willing to save the study money when they could.

3.4.7 Schedule Changes

Because the initial scheduling round had required changes in the dates for some schools--sometimes more than once--we required approval by the field supervisor (who had also managed the scheduling) for any field changes. Data collectors in general were busy enough that they needed all the time we had allocated for each visit, and they rarely requested permission to change their schedules. We granted approval for minor schedule changes in three instances.

3.5 EXPANSION OF CONTROL GROUP DATA

After the study had gone into the field, ED requested that more data be collected on the control group sample than had been planned and than the control group data collection form would accommodate. Advanced Technology agreed to collect these data by using selected questions from the Student Record Abstract, beginning with the institutions scheduled to be visited on Monday, April 18, the beginning of the second week of data collection.

We were able to telephone all 15 field staff to tell them about this change in procedure. We carefully explained which questions of the SRA should be answered for the control group students. We also sent a printed field memorandum (Figure 3-5), but this could not reach all the data collectors by April 18.

The data collectors were reporting difficulty in completing their assignments within the time allowed; the extra time required to fill out a large part of the SRA rather than the original control group form promised to prevent them from maintaining their schedules. Therefore, we made arrangements to hire and train auxiliary data collectors to make separate visits to any institutions with control group samples of five or more and complete the control group data collection there. Through the deputy project director's personal contacts, we hired a retired financial aid director and one whose school was temporarily closed. They came to the project office for three days of training which included a field practice and covered the entire

PELL GRANT QUALITY CONTROL PROJECT
SPRING, 1983 DATA COLLECTION
FIELD MEMO #3

The following items make explicit the new procedure for the Institution Control Group (ICG).

- Draw field samples as originally planned.
- Do not fill out the white, 4-page, ICG form; instead, use the blank SRAs, filling out the following items:
  - -- Q1-Q8 (all of page 2)
  - -- Q14 from the ICG (write or tape to inside of front cover facing page 2)
  - -- Q45-Q47 (bottom of page 12)
  - -- Q48-Q56 (sections II-III, pages 15-17)
  - -- Q62 to end (pages 21-36)
- Copy the student name, student study ID, and student's SSN (items 1, 3, 4) from the ICG to the front of the SRA.
- Write "ICG" in the top right corner on the front of each SRA you use in place of an ICG form.
- We will plan to send an "aide" to any school where you have 5 or more ICG students already selected. Aides will go into the field on May 2. We will try to get aides to schools with 5 or more ICGs during your visits; schools you visited between April 18 and April 29 will be visited by aides between May 2 and May 20.
- We will plan on your completing SRAs in place of ICG forms at any school where you have 4 or fewer ICGs.
- If you have time to do all the ICGs at a school with 5 or more, please do so and tell Albert Parker about it.
- Since ED may change its mind again, do not dispose of ICGs until you have completed a school; we may switch back to them. Do not send unused ICGs to the project office.

FIGURE 3-5

FIELD MEMORANDUM ON
CONTROL GROUP DATA COLLECTION

SRA, equipping them to assist the primary data collectors with the main sample when necessary.

We had a tentative schedule for these auxiliary data collectors which consisted mostly of schools which had not yet been visited by the primary data collectors. It eventually developed that the primary data collectors were able to complete all of their control group forms at many of the schools where we had expected them to need help. Evidently, they became experienced enough with the SRA to have sufficient time to undertake the added burden of the expanded control group data. In order to reduce travel time and costs and limit data collector travel fatigue, we tried to schedule the auxiliary visits to the remaining large schools on a regional basis. This required in two instances that the auxiliary data collector visit the institution **before** the primary data collector did the interview. At six institutions, the auxiliary data collector's visit was concurrent with the primary data collector's. At another four institutions, the auxiliary data collector completed the control group forms after the primary data collector had left.

3.6 FIELD PERSONNEL PROBLEMS

During the data collection period, several problems arose which required personnel reassignment and other management decisions. During the recruitment of data collectors, we had identified one fully-qualified candidate who was unable to spend a full six weeks in the field but who was available for short-term emergency situations; we made extensive use of her time.

One of the candidates whom we had offered a position declined it on the Thursday before the training week. Although we were able to recruit a replacement, he could not go into the field until the Friday of the second week of data collection due to other commitments. He participated in the training and the substitute covered his schools for the first nine working days.

In the middle of the third week, a data collector notified us that she would have to resign because of the critical illness of a family member. Our substitute took over her region for the fourth and sixth weeks of the data collection period, but was reluctant to stay in the field for three full weeks. The decreased need for the

auxiliary data collectors was evident by then; we trained one of them in interviewing as well as SRA completion and scheduled him to fill in for the fifth week of the field period.

One of our data collectors became ill herself. She was able to get home at the end of the third week and was advised by her physician not to work for a few days at least. We sent a project analyst to do the data collection at her next institution. Fortunately, she recovered and was able to resume her schedule after missing only that one institution.

3.7 DEBRIEFING

We brought all but one of the data collectors back to the project office on the Saturday after the last week of the field period for a debriefing (the absent data collector had a family emergency but submitted an extensive written report). Dr. David Iwamoto and Ms. Jean Saunders from the Department of Education observed the morning sessions, which were concerned with substantive Pell Grant issues. Seven members of the Advanced Technology professional project staff also attended.

About a week before the debriefing each data collector had received a formal debriefing agenda (Figure 3-6) which listed the topics for discussion. During regular telephone supervision calls, we asked them all to make informal, written notes of what they thought they should talk about at the debriefing. These notes would focus their attention and stimulate them to organize their thoughts, thus enabling them to use the debriefing time more efficiently. The notes would also reduce the possibility that important items would be overlooked or omitted. We collected notes from nine of the fifteen participants.

The debriefing agenda was designed to permit free discussion by the data collectors while ensuring that all important topics were covered. Each session was moderated by the project staff member most closely concerned with the topic through experience or project responsibility. Session moderators intervened when necessary to give every data collector a chance to participate and direct the discussion to agenda topics that would not otherwise have arisen.

PELL GRANT QUALITY CONTROL PROJECT
SPRING, 1983 DATA COLLECTION

DEBRIEFING AGENDA
Saturday, May 21, 1983

8:30 - 8:35 Introduction and purpose of debriefing - Carol Miller

8:35 - 9:15 *Institutional Validation - Dan Geller

- Extent of institutional validation
  - Which schools do and which don't
  - What types of students
- Exemplary procedures
- Non-exemplary procedures (what not to do)

9:15 - 9:45 *Institutional Error - Dan Geller

- Award calculation error, including payment schedule look-up
- Monitoring satisfactory academic progress and enrollment status

9:45 - 10:00 *Financial Aid Records - Dan Geller

- Completeness of records at financial aid offices
  - Paper records
  - Access to computer files maintained by other offices
- Completeness of validation documentation
  - Tax returns
  - Social Security and VA benefits
  - Other documentation

10:00 - 10:15 BREAK

10:15 - 10:45 *Disbursements - John Neely

- Frequency and method of disbursements
- Recovery of overpayments
- Accommodation of validation delays
  - Extent of payment delays
  - Methods of accommodation

10:45 - 11:15 *Financial Aid Packaging - John Neely

- General reactions by FAAs to packaging questions
- Variations in use of non-Pell aid
- Adjustments for Pell overawards and underawards

FIGURE 3-6

DEBRIEFING AGENDA

11:15 - 11:45 *Campus-Based Programs - Richard Tombaugh

- Differences between Pell and Campus-Based
- Additional observations about Campus-Based

11:45 - 12:15 LUNCH

12:15 - 12:40 **Data Collection Instruments** - John Neely

- Problem items
- Format improvements
- ICG Forms

12:40 - 1:00 **Training Program** - Richard Tombaugh

- Effectiveness of training materials:
  - Manuals
  - Exercises
  - Practice Visit
- Effectiveness of training on:
  - Conducting the interviews
  - Dealing with different kinds of records
  - Clock-hour schools
  - Difficult respondents and situations
  - Answering questions about the study

1:00 - 1:15 **Analysis of Data** - Dan Geller

- Summary of findings to date
- Quality of data coded so far

1:15 - 1:45 **Logistics** - Albert Parker

- Itineraries (hotel, air, and car arrangements), schedules
- Supervision
  - Field monitoring
  - Validation feedback
  - Call-in procedures
- Material support
  - Expense advances and paychecks
  - Field memos
  - Forms resupply

1:45 - Completion **Final Business Close-Out** - Tracy O'Connor

- Reconciliation of final expense reports
- Collection of unused tickets and materials and of car rental records
- Distribution of homeward airline tickets

*These sessions will be documented.

FIGURE 3-6 (cont'd)

DEBRIEFING AGENDA

The substantive morning sessions were taped. We compiled a transcript of these sessions to augment the written notes submitted by the data collectors and to use as an additional source in program analysis. A project staff member also took extensive written notes during the entire debriefing. The morning debriefing sessions were the subject of a report, which included the transcript of all the taped sessions.*

The afternoon debriefing sessions were devoted to topics relating to the administration of the data collection--instrumentation, training, and logistics--and to final business arrangements with the data collectors. The afternoon sessions will provided a basis for improvement of future data collection efforts.

*Advanced Technology, Inc., "Report on Debriefing of Spring Data Collectors for Stage Three of the Pell Grant Quality Control Study," June, 1983.

4

STUDENT AND PARENT DATA COLLECTION

Westat collected data during the Spring of 1983 from a nationally representative sample of 4,109 Pell Grant recipients and their parents. The sample of students was selected from 317 participating institutions.

This chapter discusses the following aspects of the field data collection effort:

- Organization of the field force
- Supervisory and interviewer training
- Field operations
- The Automated Survey Control System (ASCS)

4.1 ORGANIZATION OF THE FIELD FORCE

To manage the Pell Grant interviewing staff and to coordinate field operations, Westat divided the 48 states into 7 supervisory regions (see Figure 4-1). Five of the seven supervisors had also been supervisors on the previous round of Pell Grant (formerly BEOG) data collection in 1981. The other two supervisors had extensive interviewing and supervisory experience on other, similar studies. The supervisors reported directly to the Westat field director, who in turn, reported to the Westat project director.

Regional supervisors and their assistants began recruiting interviewers in December, 1982. Primary sources for recruitment included Westat's computerized interviewer file, including the file of previous BEOG interviewers; supervisors' local contacts; local employment agencies; and, when necessary, newspaper advertisements. The location of sampled institutions and the number (but not necessarily the location) of students and parents associated with each institution were known at the time of recruitment. Westat used this information as a basis for identifying desirable

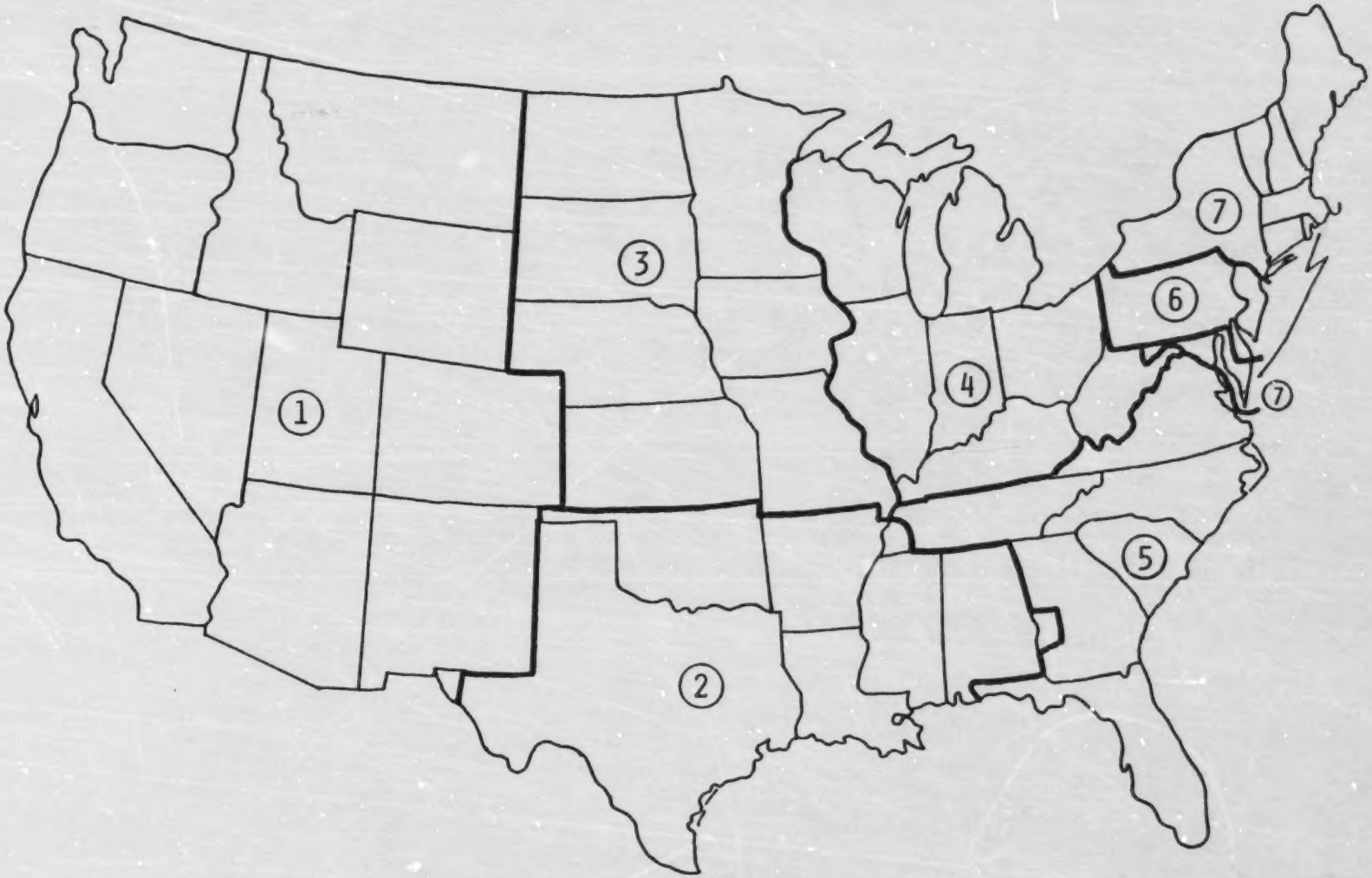


FIGURE 4-1

WESTAT SUPERVISORY REGIONS FOR
STUDENT-PARENT INTERVIEWING

locations for interviews. When interviewer training began on February 14, 1983, 142 interviewers had been recruited and invited to training.

4.2 SUPERVISOR AND INTERVIEWER TRAINING

Westat carried out a training program at its home office for regional supervisors. Training sessions for individual interviewers were held at four sites throughout the United States.

The five supervisors who had worked on the previous round of the Pell Grant Study were sent the study materials and the revised Supervisor's Manual. They were instructed to review and study these materials and discuss any questions or problems with the field director.

The two new supervisors came to Westat's corporate headquarters for training on February 7-9, 1983. They also were sent study materials to review prior to coming for training. At the training, project staff reviewed the history and purposes of the study, discussed all study materials, and presented the responsibilities of the supervisor. An important part of training was instructing the supervisors on how to use the ASCS and giving them time to practice with the system. The final day was spent with the Westat supervisor in Baltimore. This field trip was organized to give the new supervisors the opportunity to see how an experienced supervisor organized her office for the Pell Grant study and to discuss with her the problems and details of the supervisor's job.

The goal of Westat's interviewer training sessions was to develop and hone the interviewing skills of the potential interviewers to enable them to collect survey data in a uniform and professional manner. Interviewers were instructed in three areas: general interviewing techniques, field procedures specific to this study, and an understanding of the intent and use of the survey instruments. The greatest amount of training time was devoted to questionnaire-related issues, such as presentations of question specifications and subsequent role playing of practice interviews, for which mock respondents used a scripted answer guide.

A total of 141 trainees were trained in one of four sessions, each lasting four days. During the first week (February 15-18), training was held in Philadelphia and St.

Louis; during the second week (February 22-25), training was held in Nashville and San Diego. Two training teams led these sessions--both teams having prepared their presentations and strategies together during the week of February 7. This joint preparation, combined with the use of a scripted Trainer's Manual, helped to maximize consistency within and between these two teams.

Westat based interviewer training for the Pell Grant survey on a training plan which the company had developed and refined for other large-scale surveys. Training techniques included home study of a programmed text, interactive lectures to the entire group, and role playing (administration of a mock interview complete with documentation). Central to the training plan was the subdivision of the group of trainees into small groups of approximately 10 trainees who received individual attention from the group leader. The field supervisors functioned as group leaders during training so that they were able to observe and assist the interviewers they would later supervise. Additional trainers also acted as group leaders and reported the progress of each trainee to the supervisors at the end of each day of training. Trainees who successfully completed a practice interview, the final exam, and who received positive evaluations from the group leader were given their initial assignments at training.

4.3 FIELD OPERATIONS

In general, the field work went very smoothly. Supervisors had scheduled conferences with each interviewer at least once a week, or more frequently if necessary. Similarly, the supervisors were scheduled to report to the field director at least weekly, although interaction between the home office and supervisors occurred almost daily.

Assignments to interviewers were made by each supervisor on an "as needed" basis. Most interviews received new assignments weekly. The supervisors monitored the number of assignments being worked on by each interviewer so that no interviewer had a backlog of more than a week's work. Interviewers were encouraged to schedule their appointments so that each week they would work a minimum of 20 hours. They were also encouraged to schedule student interviews first, since the students were much more likely to become unavailable as the field period extended into May, when many institutions end their spring terms and begin summer vacation.

In many cases, the interviewers had reasonably current addresses for the students that they were assigned. Most were easily tracked through student directories or the local phone book. If a student could not be located through conventional means, the interviewer referred the assignment back to the supervisor. The superivsor then called the appropriate office in the institution, such as the Student Affairs Office, to acquire a current address.

In scheduling interviews with students, interviewers occasionally were told by a student that they had never received a Pell Grant. This is not surprising since students were sampled from both the recipient file at the school and the pending file of students who had been determined to be eligible but who had not yet received a grant. Such cases were referred to the supervisor, who called the financial aid administrator at the school to verify that the student was not a Pell Grant recipient.

Up to five telephone and three in-person attempts to obtain an appointment with each student or parent were required. Most appointments, however, were successfully made in fewer contact attempts.

Although student respondents were generally located near the sampled institution, parent respondents were scattered throughout the entire country. This meant that at some point, some interviewers had to travel to respondents located in far away places. Assignments of this type were held and allowed to accumulate until the last few weeks of the field period. At that point, supervisors coordinated interviewer travel plans with the field director, and the interviewers were sent out to conduct the interviews. Because of the dispersion of these parent interviews, they were often conducted by an interviewer different than the one who conducted the corresponding student interivew.

This procedure was not followed for 261 parents of independent students who lived distant from any interviewer. Documents for these interviews were shipped to the Westat Telephone Center, where the interviews were completed. Since the interview with such parents was very short, we considered this to be a satisfactory way to reduce interviewing costs without jeopardizing the quality of the data collected.

4.4 THE AUTOMATED SURVEY CONTROL SYSTEM (ASCS)

An integral part of field management on the Pell Grant Study was the computer-assisted management system known as ASCS. The ASCS operated through small computer terminals located in the supervisors' homes and connected through telephone lines to a computer. A similar terminal was located in Westat's home office for use by the field director. Each week the supervisors would enter information on field progress into the system, and on a regular basis the terminal would print out summary reports on survey progress. The system was also used to transmit and receive messages to and from the home office, as well as among the regional supervisors.

The ASCS generated three reports which were used by the Regional Supervisors. ASCS Report #1, the Supervisor Interview Report, listed the identification numbers of all cases currently assigned to an interviewer. Each week the system generated a new Report #1 for each interviewer. This report served as a record of assignments and was discussed during the weekly supervisor/interviewer conference. ASCS Report #4 also was generated weekly and presented production information on all interviewers. It provided information such as response rate, hours and expenses per completed interview, and cost per completed interview. ASCS Report #5 presented totals showing the current disposition of all cases in a region. A review of this report provided an accurate, overall picture of a region's progress in completing the survey.

In addition to these reports, the ASCS was capable of generating 14 more reports for use by home office staff. These reports were used to monitor nationwide survey progress and provided detailed information about different respondent types (e.g., dependent students or parents of dependent students).

4.5 SECONDARY DATA COLLECTION

Some of the items on the Pell Grant application that was filled out by each of the students in the sample were verified using an independent source. The purpose of acquiring secondary data was to verify income and asset information with "harder" documentation than the self-reported income and asset data provided by the student or parent during the interview. The major source of secondary data, used to verify income variables, was the Internal Revenue Service. Other sources were Financial

Institution Records (FIR), used to verify amounts deposited in savings and checking accounts; and Tax Assessor Records (TAR) used to verify the home value for a 25 percent subsample of homeowners.

The secondary data collection lagged behind the field interviewing by several weeks, because the necessity of acquiring additional data was partly determined through each student or parent interview.

4.5.1 IRS Tax Returns

Westat made arrangements with IRS to acquire copies of each respondent's tax returns for the 1981 tax year (for those who filed a 1040 or 1040A) or a notification from the IRS that a particular respondent's 1981 tax return was not found. If the IRS could not find a return, we assumed that the respondent probably did not file a 1981 return.

The procedure specified by the IRS required that a release form be signed by each taxpayer authorizing Westat to obtain copies of the tax payer's 1981 return. IRS Form 4506 is the release form used for that purpose. Westat used a specially modified version of Form 4506 for the Pell Grant project.

IRS Forms 4506 were sent to each of the students and parents in the sample prior to the interview, with instructions to fill out the form and return it to Westat. At the end of the interview, the interviewer asked the respondent to sign another Form 4506. While this double procedure produced many duplicate release forms, it ensured more complete coverage than might be expected by using a single method.

Westat sent 6,145 original Forms 4506 to the IRS service centers, which returned 6,032 to Westat. Of the 6,032 returned, usable copies of tax returns were obtained for 4,007 survey respondents, the remainder (2,025) represented presumed non-filers for whom no returns could be found.

4.5.2 Financial Institution Records (FIR)

During the in-person interview, respondents who reported having more than $4,000 in checking or savings acounts at the time of application for a Pell Grant were asked to sign a release form for each account so that Westat could obtain the balance on those accounts directly. (A small number of respondents who did not know the value of their accounts were also asked to sign releases.) As the releases were received at Westat, they were logged in and the date of application for each release was added to the forms. They were then sent to the financial institution with a cover letter explaining the study and a self-addressed envelope. Considerable telephone follow-up was required to persuade institutions to reply. Of the 392 releases that were sent to financial institutions, 365 were returned to Westat.

4.5.3 Tax Assessor Records (TAR)

For a 25 percent sample of homeowners (among parents of dependent students and independent students who reported owning homes at the time of application), the local tax asessor's office was contacted by mail to acquire home value. As with the FIRs, considerable telephone followup was required, both to identify the appropriate jurisdiction and to encourage a reply. Westat sent 365 forms to tax assessors; 364 were returned. However, since some of these 364 homeowners were later found to be ineligibile (either the student did not receive a grant or did not attend the school where sampled), 349 TARs were included in the data file.

5
DATA PROCESSING

This chapter explains not only the automated data processing procedures used at Advanced Technology and Westat, but also the premachine processing, that is, manual coding and editing. Quality control procedures for the required software are also discussed.

5.1 INTRODUCTION

Data for the project were derived from nine different sources, with different record and file structures and requiring different amounts and levels of processing to be used for analysis and production of the final report. The nine data sources were:

- Student Questionnaires (SQ)
- Parent Questionnaires (PQ)
- Student Record Abstracts (SRA)
- Institutional Interview Forms or Questionnaires (IQ)
- Institution Control Group Forms (ICG)
- Computed Applicant Records of Student Aid Report (SAR)
- Tax Assessor Records (TAR)
- Financial Institution Records (FIR)
- Internal Revenue Service 1040/1040A Forms

Westat, Inc., administered the Student and Parent Questionnaires, collected the TAR, FIR, and IRS data, and cleaned and edited all the data from these sources. Advanced Technology received both Westat's data and Westat's edit specifications. The purpose of these reviews was to assure that Westat quality control procedures were followed and were effective. Advanced Technology monitored Westat's data

collection schedule closely to ensure that no schedule delays were created that might have affected Advanced Technology's processing of the data. Upon receipt of data, a series of frequencies and descriptive statistics were run in order to ensure data accuracy. This procedure served as a final step in the data editing process and provided assistance in the following areas:

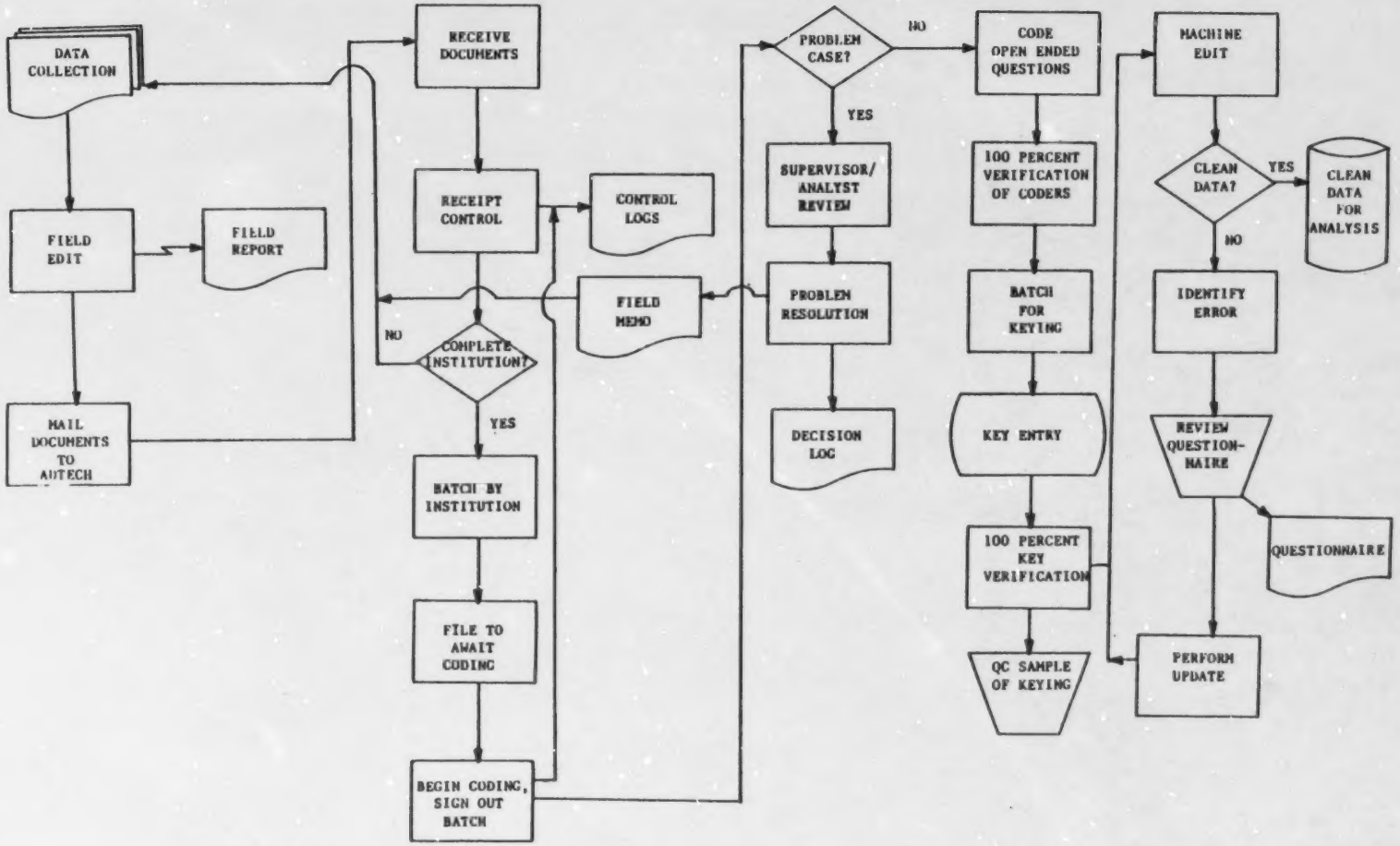
- Identification of missed edit checks
- Identification of missed data updates
- Preparation of the final tabulations plan

Advanced Technology prepared the specifications for frequencies and descriptive statistics using SAS. The remaining data (i.e., SRA, IQ, ICG) were processed in line with the procedures outlined in this report.

The plan addressed six groups of procedures:

- Data control procedures
- Coding procedures
- Data entry procedures
- Edit and update procedures
- SAS merge procedures
- Documentation procedures

Each procedures group was composed of several steps or tasks which, when combined, produced a clean file data set for analysis. This process is depicted in the following figures. Figure 5-1, "Pell Grant Quality Control Project--Stage Three Data Processing Plan Data Flow" is an overview of the flow of data from collection through the production of a clean data file for analysis. Figure 5-2 "Pell Grant Quality Control Project--Merge and Analysis" shows how these data were combined to create a single file for analysis.



70 FIELD CODING AND EDITING DATA PROCESSING 71

FIGURE 5-1

PELL GRANT QUALITY CONTROL PROJECT--STAGE THREE
DATA PROCESSING PLAN DATA FLOW

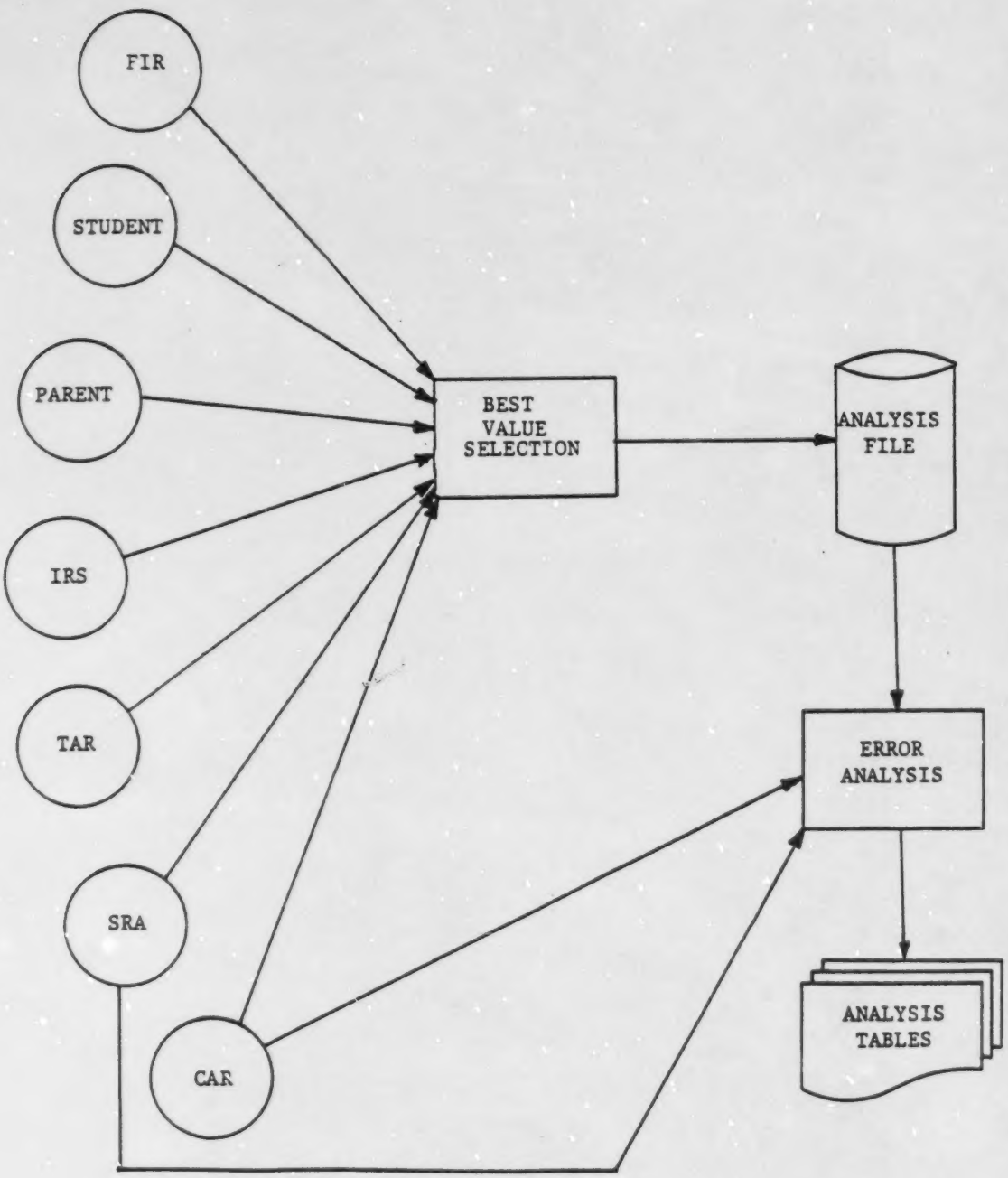


FIGURE 5-2

PELL GRANT QUALITY CONTROL PROJECT--MERGE AND ANALYSIS

5.2 INSTITUTIONAL DATA

Advanced Technology collected two kinds of data during the institutional visits. Data about the institution and its financial aid administration procedures and packaging philosophy were the subject of the Institutional Questionnaire. The site visitors also examined the institution's records on each student and completed an individual Student Record Abstract. Site visitors mailed the IQ and associated SRAs to Advanced Technology's project office shortly after the end of each visit (after field editing).

5.2.1 Data Control Procedures

Data control is a critical part of any data collection effort. In this particular application it became especially important because there were multiple sources of data dealing with the same respondent. A tightly controlled tracking instrument was necessary to monitor the processing status of questionnaires and prevent duplicate entry of data. For these purposes, a series of data control logs were established which indicated:

- Location of an institution
- Interviewer responsible for an institution
- Status of documents for a given institution
- Batch number for each institution
- Students within a given institution

Logs for the SRA and ICG consisted of three levels of control. First was the Student Listing Log (Figure 5-3), a computer-generated listing of all sampled students within each institution. As forms came in from the field, they were checked off against this log and any problems associated with either the institution or individual students were recorded. The Institution Control Sheet was the second level of control (Figure 5-4). This form acted as a control form for institutions and provided summary statistics for each institution (e.g., number of documents for each institution sampled) and provided tracking of data received, coded, and keypunched. The third level of control was the Batch Control Sheets (Figure 5-5), which were attached to the front of

SPRING 1983 DATA COLLECTION : STUDENT LISTING

INDIANA UNIVERSITY BLOOMINGTON
BLOOMINGTON IN
INSTITUTION NUMBER 17

	SSN	STUDENT	STUDY_ID
1	304-52-9741	[redacted]	17001
2	314-60-6413	[redacted]	17002
3	313-80-7776	[redacted]	17003
4	308-76-7081	[redacted]	17004
5	314-78-6534	[redacted]	17005
6	586-58-8607	[redacted]	17006
7	317-70-6920	[redacted]	17007
8	312-48-8516	[redacted]	17008
9	231-11-4177	[redacted]	17009

FIGURE 5-3

STUDENT LISTING LOG

INSTITUTION CONTROL SHEET

I. RECEIPT INFORMATION:

DATE RECEIVED _____ INSTITUTION ID _____

DATA COLLECTOR _____

NUMBER OF SRAs _____

NUMBER OF CCGs _____

IQ RECEIVED?

II. CODING INFORMATION:

	Coder	Date	Reviewed
SRAs			
CCGs			
IQ			

III. KEYPUNCH INFORMATION:

	Sent	Returned	Batch
SRAs			
CCGs			
IQ			

COMMENTS:

FIGURE 5-4

INSTITUTION CONTROL SHEET

Batch Number	Sent	Returned	Form Type

		ID	Coder	Date	Number
INSTITUTION	(1)				
	(2)				
	(3)				
	(4)				
	(5)				
	(6)				

TOTAL _____

FIGURE 5-5

BATCH CONTROL SHEET

each keying batch to maintain counts and responsibility for assignments throughout the keying, editing, and updating. A copy of the Batch Control Sheet was also kept in a looseleaf notebook so that the coding supervisor could track forms from a central location at all times.

The Student Listing Logs served both the SRAs and the ICGs. Therefore, there were no institutional cover sheets for batches of institutional questionnaires.

All documents were received and batched according to institution. When the data preparation staff received a package from a sampled institution, they divided the contents into groups by form type (i.e., SRA, IQ, ICG). The coding supervisor recorded all forms in their respective logs and determined coding assignments. When the volume of forms received for one day exceeded the coding capacity of the staff, forms were held in a secure file to await coding assignment. After coding and all appropriate logging procedures, forms were placed into separate files to be batched for data entry.

Before going to data entry, forms for institutions were grouped into keying batches. These batches then served as the quality control group throughout keying, editing, and updating. Each batch was assigned a number and a Batch Control Sheet (Figure 5-5). A Batch Control Sheet was a listing of all institutions in a batch and included the information in the log book. These forms and procedures allowed precise tracking of all survey instruments from receipt through coding, logging, editing, updating, and data retrieval.

The construction of a Survey Control File also contributed to the plan for data control. Survey Control Files were developed for both the Institutional Questionnaire and the Student Record Abstract files. These files consisted of the identification number, the region code, and dependency status code for each corresponding record in the various surveys. These files served as a further check on data accuracy by assuring that all ID numbers and accompanying identification information (i.e., region, dependency status) were properly keyed and that no duplication or extra cases occurred. This was accomplished by performing a merge of the newly keyed data and the respective Survey Control File. Any discrepancies in this merge process were flagged and resolved.

5.2.2 Coding Procedures

Manual coding and cursory editing were done under the close supervision of the coding supervisor. The coding supervisor's responsibilities included:

- Logging incoming forms
- Making coding assignments
- Monitoring coding production and regular reporting to the project director
- Maintaining all logs and tracking forms
- Assisting in error resolution
- Assisting in coder training
- Maintaining lists of codes for open-ended questions

The coding supervisor and the analysis staff conducted coder training. Training consisted of an introduction to the study, the data collection instruments, and standard coding conventions and procedures. After the introduction, coders were instructed on the mechanics of coding every question on each of the forms (SRA, IQ, ICG). Coder training also included step-by-step instructions in the keeping of logs, the procedures for resolving problems in the data, and the system for updating information.

The Coding Manual and Data Elements Dictionary contained comprehensive information about all possible codes, acceptable ranges, and the coding conventions and procedures being used for the data collection. Each coder had a copy of the manual and was responsible for making all of the necessary updates to his copy. Any changes made to the coding manual were made during the daily meeting of the coding staff. Since the manual was the primary reference for any questions of coding and code definitions, it was imperative that each coder's manual be kept up to date and accurate. Any changes had to be added to the manual, dated, and initialed by the coding supervisor. The coding staff met daily to update the data elements dictionary and to ensure that new procedures and codes were properly implemented.

During the training sessions all coders were instructed in the coding and editing procedures for use with each of the data collection instruments; however two coders

were assigned primarily to the coding and editing of the IQ and the ICG. Since all of the coders were trained to code all three forms, coder time was reallocated as needed to meet changes in workload or time pressures that developed during the course of the data collection.

Problems encountered during the coding process were documented and resolved. The coders brought all problems of inconsistency within a questionnaire, responses out of acceptable ranges, illegible responses, need for new codes, or incomplete responses to the attention of the coding supervisor. The form in question was set aside until the problem had been resolved. When problems arose with the data that required a higher authority for resolution, the coding supervisor brought them to the attention of the manager of data analysis or the technical director.

The standardized coding conventions employed on all three forms inlcuded the following:

- All numeric entries were right-justified and zero-filled.
- A code of "99" was used to indicate that a question was asked but no response was given.
- A code of "97" always indicated that a question was asked but the respondent was uncertain or could not determine the answer.
- Blank spaces on the form indicated that a question was not asked due to a skip pattern.
- All changes, corrections, and updates were made in red ink.

All forms received were opened, logged in, and filed in a secured room. All coding and editing were done in an office adjacent to the file room. Restricting the receipt of forms, coding operations, and filing to two rooms limited the possibility that any data would be lost or misplaced and assured the security and confidentiality of the data.

In addition to the detailed coding specifications described above, we devised a graphic representation of all data records in the form of data record layouts to accompany each codebook. These record layouts provide a graphic representation of the physical position of all of the data elements in a given record as well as logical

groupings of the various data elements. This graphic presentation aided in the development of the comprehensive logic checks for the editing process. The data definitions for all programs accessing these data were derived from record layouts. In the event that OSFA ever requires a documented OS file, record layouts will prove an invaluable resource in preparing documentation.

5.2.3 Data Entry Procedures

During the production phase of data collection, entry and editing, we sampled a small group of questionnaires to serve as a quality control check on the data entry vendor. On this sample we compared keyed data from data entry with hardcopy documents to estimate the keying error rate. Errors could be the result of misinterpretation of the interviewer's marks and might not be caught during verification. This sample quality control check found no data entry errors by the data entry vendor.

We described all edit checks in a Computer Edit Instruction Manual. This manual had two purposes: 1) to provide detailed specification to the programmer for the preparation of the edit programs; 2) to serve as a tool in error resolution prior to the update cycle. Based on the coding scheme developed for the various codebooks, a comprehensive series of range checks was applied to all variables in each survey instrument. Range checks tested for all valid codes and for missing codes.

We employed a consistent set of coding conventions for all the various survey instruments to aid in coding and the resolution of errors. Missing value codes were the same for all questionnaires and, whenever possible, we assigned the same code to applicable valid responses which could be given to several questions.

To identify all applicable logic tests, we developed a data flow diagram for each survey instrument. This flow chart illustrated the various skip patterns that appeared in each questionnaire and aided in identification of significant relationships between questions. In addition to those skip pattern checks, we made checks for all apparently inconsistent responses. Inconsistent or contradictory information in the data was detected by the logic check module and was reviewed by an analyst.

The coding supervisor and systems analyst developed range and logic tests that were reviewed by the manager for data analysis. After a draft of all checks had been prepared, the entire data processing team performed a structured walkthrough of them. This walkthrough helped to identify any additional checks needed to eliminate redundancy, and to minimize the possibility of error from the interviewers, coders, or the data entry vendor.

The edit programs produced error reports which were forwarded to the coding supervisor for error resolution. The coding supervisor, using the Computer Edit Instruction Manual, compared error reports with original questionnaires and resolved the problem using the predetermined error resolution techniques. The resolved edit report was forwarded to coders to prepare an update transaction which was keyed into an ALPHA transaction file. This transaction file was processed by the update program, which posted it against the master record. The update program also produced a report detailing all transaction records, additions, deletions, and modifications (displaying the record both before and after modification) with summary statistics. The data processing procedure included the use of a batch update program. The update program, like the edit program, was written in COBOL for ease of use and maintenance. There were several distinct advantages to using a batch process for the update cycle:

- Greater quality control was achieved since each update transaction had a printed record for auditing.
- Coders spent less time on line. They were able to enter transaction records without searching the files.
- Each update pass wrote a new file so that the old version with the transaction file served as a backup.
- Access to the raw data was limited to the batch cycle under the control of the coding supervisor, assuring greater data security.

The file management structure required that static batches (those not yet pronounced "clean") remain on disk as members of a partitioned data set (PDS) to be updated by the batch cycle. As each batch was pronounced clean, it was transferred to tape to await concatenation with other clean batches. This provided enough efficiency to allow the use of traditionally scarce disk space at COMNET.

Backups consisted of old members of the static PDS. In the event that they could not be used, the original data transmitted from the data entry vendor were available. Since there was an audit trail of update transactions, it was a relatively simple matter to reconstruct the most current data from the original.

5.2.4 SAS Merge Procedures

After the data had been edited and appropriate updates had been made, the data (SRA, IQ) were combined with other data from Westat (Parent, Student, IRS, TAR, and FIR Data) to create the analysis file. This file was created through a complex series of merges performed in SAS. Westat prepared a data element dictionary of the Stage One files, which the data processing and analysis staff reviewed. However, the staff decided that new naming conventions would be more useful.

To produce an accurate analysis file using new code, the programming team and the analysts reviewed the "best value" selection process to develop a more efficient means of producing the analysis file. To aid in this review process, a new systems flow chart detailing the Stage One merge was developed. This flow chart helped the team identify points at which efficiency could be enhanced.

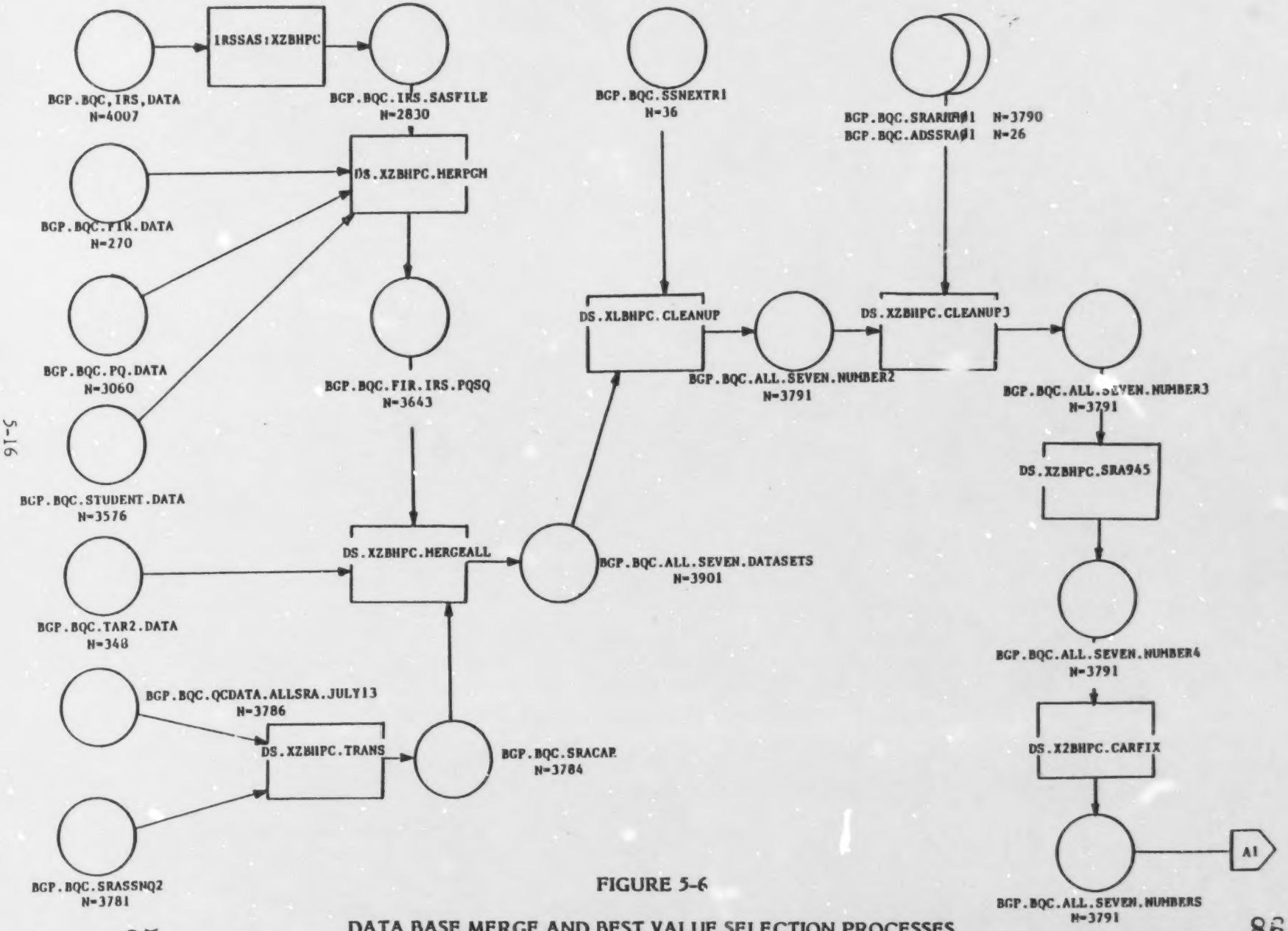
The use of SAS for the merging of the Pell Grant data offered significant advantages because of power and flexibility. However, caution had to be employed when merging datasets to insure that the desired values were incorporated. The merging of multiple files with many shared, similar, and discrete variables is vulnerable to error in several ways. Files do not always merge exactly as planned. Sometimes variables, or even entire records, are lost without a trace because of insufficient verification of merged datasets. For this reason, reports were produced to present in detail those cases that did and did not merge in order to verify the accuracy of the software. SAS has excellent self-documenting abilities in the input and output process and provides detailed information as to what data entered a step, what data exited a step, and what happened to the data in the process. The programming and analysis staff monitored this trail of information closely at crucial points to ensure the correct execution of the program as intended.

The creation of the analysis was logically divided into two procedures: 1) a range process where all sources of data were brought together on the basis of common keys; and 2) the selection of the "best" documented value from available data. This process is depicted in Figure 5-6.

The first step in the merge process was to reformat the IRS data. Because of its hierarchical file structure, the IRS data could not be used without reformatting. Therefore, we created a flat file containing the specific variables needed for analysis. This reformatted IRS file was then combined with data from the Student Questionnaire (SQ), Parent Questionnaire (PQ), and Financial Institution Records (FIR). Since Westat had collected all of these data, the merge was performed by Westat identification code.

Our next step was to combine the data from the Student Record Abstracts (SRA) with the data from the file of Computed Applicant Records (CAR). Because the CAR file did not contain Westat's identification code, we performed this match using the student's social security number. The output of that merge (BGP.BQC.SRACAR) was then merged with the output of the previous merge (BGP.BQC.FIR.IRS.PQSQ) along with the Tax Assessor Records (TAR) to create the first composite file (BGP.BQC.ALL.SEVEN.DATASETS). We executed two additional programs after the creation of this composite file and before the best values selection. The first of these programs updated the file with SAR data on ADS applicants. The second program added information from the Pell Recipient History Master File on Section 3 of the SAR for institutions that exercised their option to submit these data on tape. We also performed several updates on the composite file to assure clean and accurate data. These began with BGP.BQC.ALL.SEVEN.NUMBER2 and ended with BGP.BQC.ALL.SEVEN.NUMBERS. This final composite file consisted of 3,791 cases that were used for best value selection.

The first variables that were computed for best value selection were those that were common to all students, regardless of their dependency status. A major purpose of these, which is explained in more detail in Volume 1, was to determine the "best" dependency status (BGP.BQC.INDDEP.BEST01). If the best dependency status was "independent" we executed the program to determine best values on all application items for independent students (BGP.BQC.INDVARS.BEST01). If the best dependency



5-16

FIGURE 5-6

DATA BASE MERGE AND BEST VALUE SELECTION PROCESSES

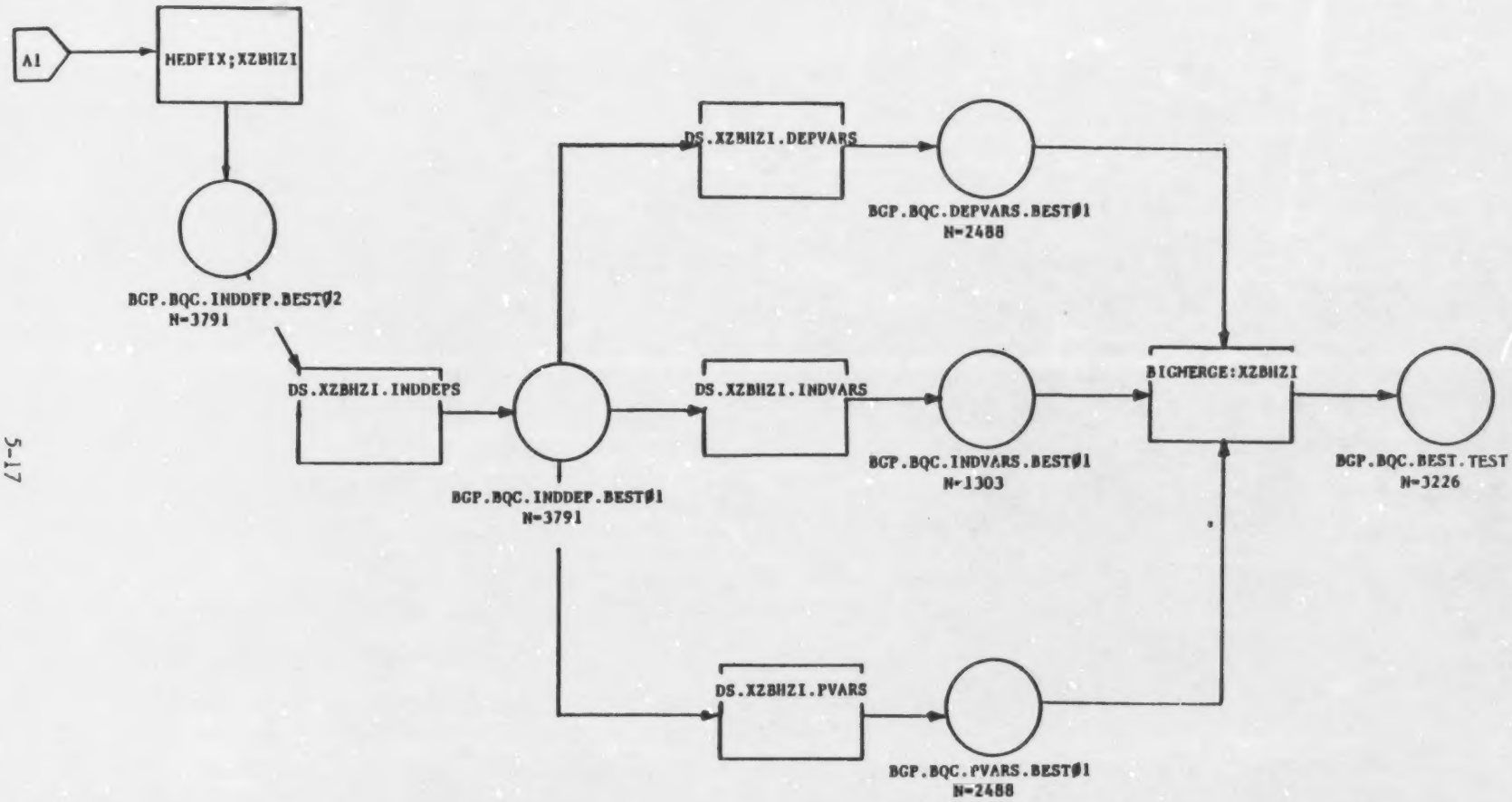


FIGURE 5-6 (Continued)

DATA BASE MERGE AND BEST VALUE SELECTION PROCESSES

status was "dependent" we executed two programs: one to determine best values for parental application items (BGP.BQC.PVARS.BEST01) and one to determine best values for dependent student application items (BGP.BQC.DEPVARS.BEST01). The merger of these best values files was a composite analysis file (BGP.BQC.BEST.TEST) used in the analysis of award errors in Volume 1.

5.2.5 Documentation

A Dataset Documentation Form (Figure 5-7) was filled out for all files to provide a more complete documentation of data files and program files. This form included detailed physical and logical characteristics of all files, whether they were OS or SAS, and whether they resided on disk or tape. Complete listings of all datasets were maintained in a looseleaf notebook. These forms provided a detailed tracking mechanism and aided in the final documentation of datasets.

SAS merge procedures were also documented internally and externally throughout the process. External documentation consisted of a detailed description of the merge cycle to be recorded on a form, one copy being bound with the computer output, the other copy being kept in a looseleaf notebook. This documentation provided an opportunity to describe in detail any problems encountered and provided the foundation of the final report detailing the merge process. The internal documentation consisted of a brief description of the inputs, outputs, and process of each SAS program and appeared at the top of the source listing as a section of comments.

5.2.6 Statistical Analysis

After editing and updating the data and creating the analysis file, we began the statistical analysis. The production of all tables was programmed using SAS, based on specifications provided by the manager of data analysis, who coordinated all analytic activities and was the primary source of requests for production of tables. The specifications for analysis included the following:

- Purpose (place in report and associated research question)
- Input files

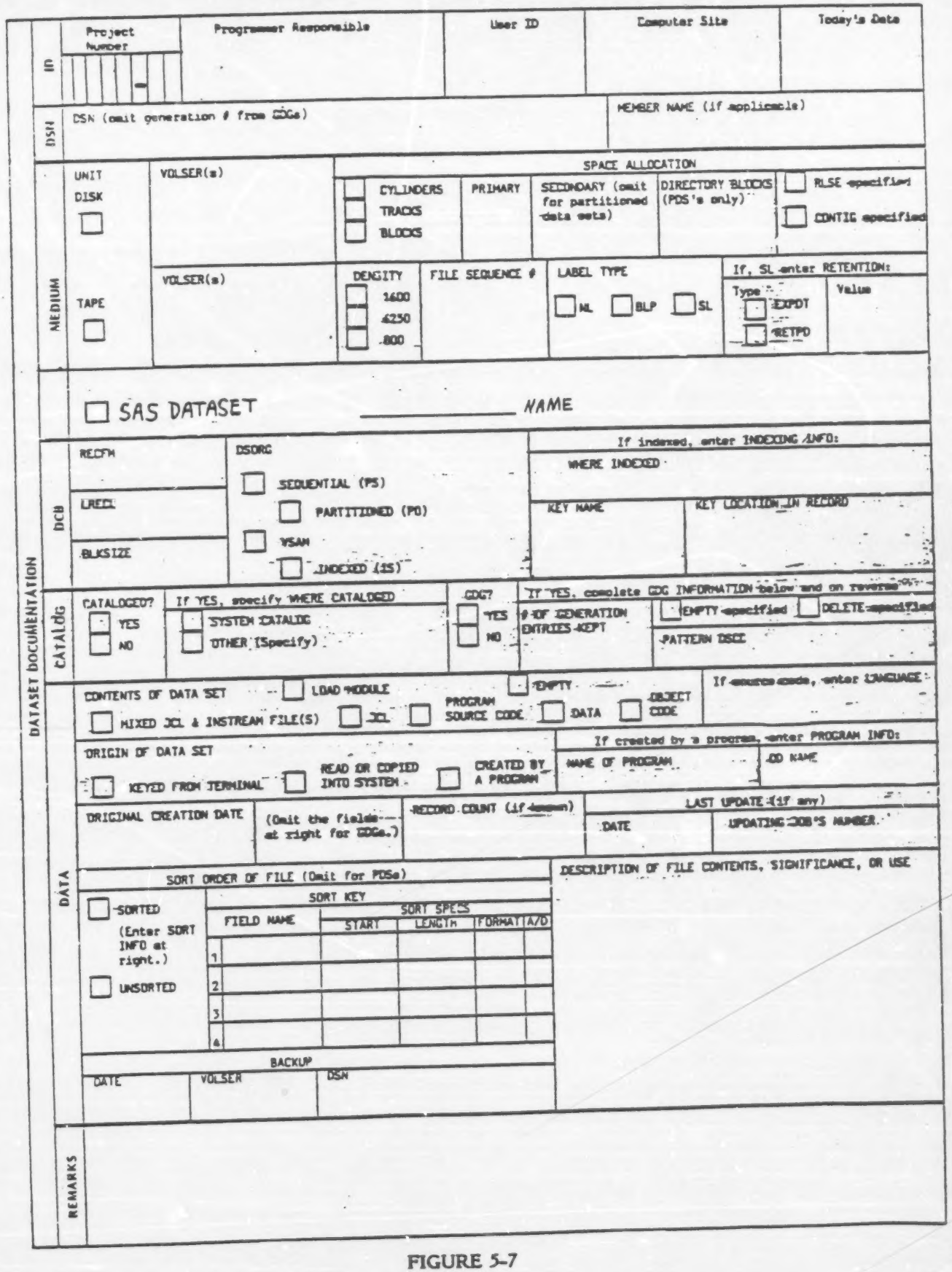


FIGURE 5-7

DATASET DOCUMENTATION FORM

- Population (which cases are included or excluded)
- Analytic specifications
- Output configuration
- Quality control output (tables designed to serve as a check on the programming)
- Output files and retention

5.2.7 Software Quality Control

Assuring the quality of software must be a prime consideration of any data processing effort. In order to maintain its commitment to the quality of the software, Advanced Technology adhered strictly to the following procedures for program development for the Pell Grant Quality Control Study:

- Structured walkthrough of edit program specification
- Structured walkthrough of merge and "best value" selection process
- Structured walkthrough of award programs
- Review of ad hoc analysis specifications by the manager of data processing and quality control analyst
- Manually produced test data to verify edit, update, merge, and award programs
- Ten percent sample of live data to test programs
- Production of marginal tables for all variables

The flow of this process can be viewed in Figure 5-8, Program Development. Once specifications for a program were developed, the staff member responsible reviewed those specifications with the quality control team consisting of the manager of data analysis, the manager of data processing, and the quality control analyst. The team determined whether the specifications met the following qualifications:

- Accuracy--are the appropriate inputs specified?
- Completeness--are all necessary variables, transformations, and subsets specified?

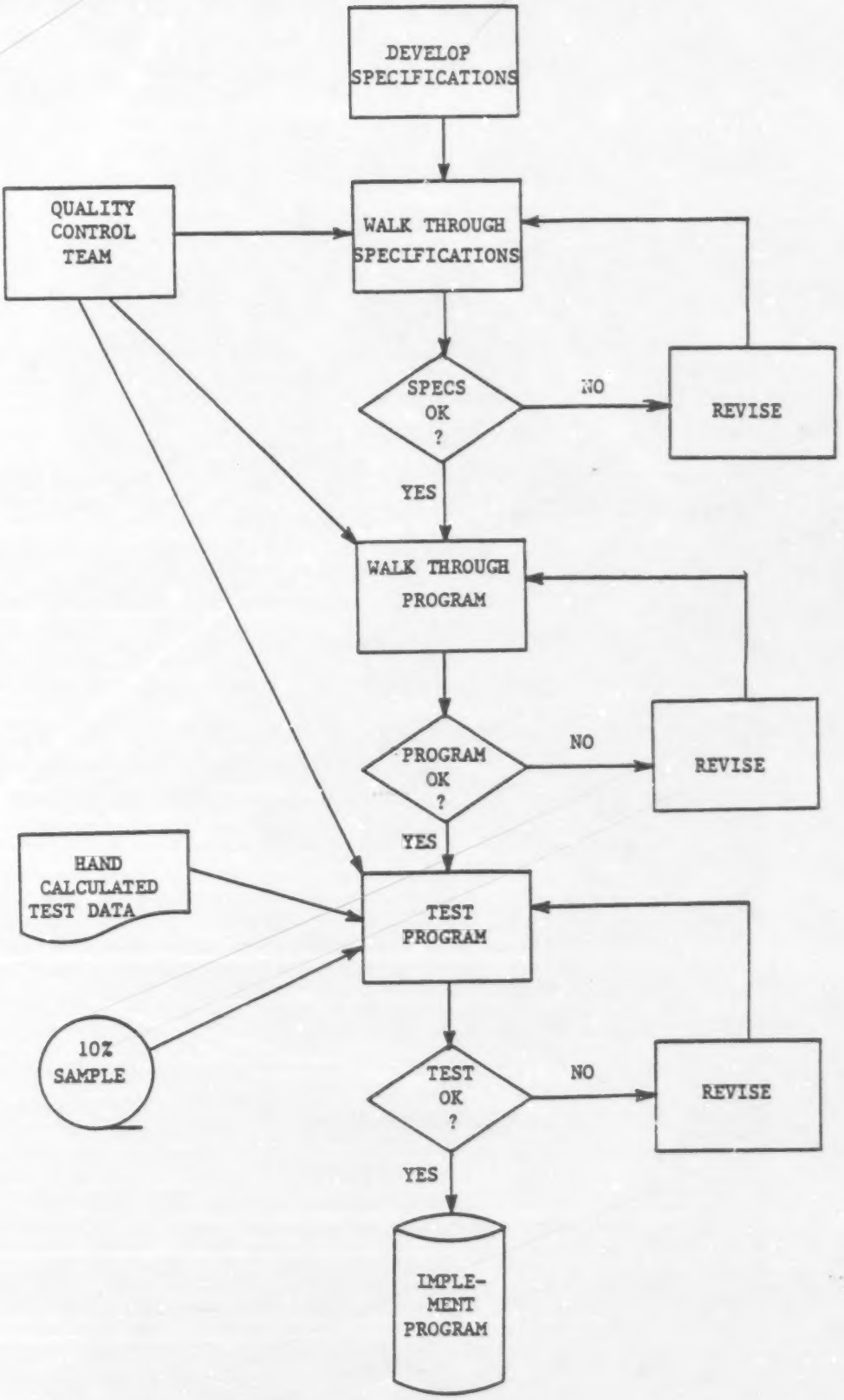


FIGURE 5-8

PROGRAM DEVELOPMENT

- Efficiency--are all the specifications necessary?

If the specifications passed these tests, they were assigned to a programmer for development. Once specifications were approved, and a program was developed, a similar process was followed to ensure proper programming.

The program was then thoroughly tested by first applying it to a batch of hand-calculated test data. The resulting output was compared to hand-calculated results to verify accuracy. The second step in testing programs was to run them on a 10 percent sample of live data. This test performed an additional check on the program to see that probable results were obtained (i.e., whether the resulting analysis fell within pre-established acceptable limits). In each of these test steps, if errors were detected, they were corrected promptly and re-tested until the program was judged acceptable.

Some data analysis was performed on an ad hoc basis. Such analysis usually required quick response and received a technical review by the manager of data processing and the quality control analyst. These reviews were less formal than the reviews performed on the edit or merge programs, but no less thorough. They examined required inputs, desired outputs, and necessary manipulation of data. Before implementation, all ad hoc programs were also put through the same testing procedures used on production programs.

The testing process was designed to follow all possible logic paths and test all boundary conditions of the software. In order to achieve this, Advanced Technology produced a package of test data that included both accurate and inaccurate data.

5.3 INDIVIDUAL DATA

Westat was responsible for the collection and processing of all individual-level data except for Student Record Abstracts, including student and parent interviews and data about individual financial status from tax assessors, financial institutions, and IRS.

The processing for each of the five independently constructed data files associated with the survey (Parent Questionnaire, Student Questionnaire, IRS Returns,

Financial Institution Records, and Tax Assessor Records) was done using very similar procedures. Data processing for the student and parent data was, however, considerably more complex and time-consuming. The work required at each step in processing the student/parent data was commensurately more complex and time-consuming.

5.3.1 Receipt Control

Receipt control, using computer-generated logs of all materials acquired relating to the student/parent sample, was the control and monitoring point for all materials sent to the field and returned to Westat. The master receipt control log listed the study identification number, name, address, and Social Security number of each student/parent pair. The log was organized in numeric order by student identifier within school. Space was provided for recording updated information on names and addresses, the date each questionnaire was received from the field, the completion status of the document, and the coding batch number assigned to the document. Labeled columns were also provided for recording the receipt of IRS Release Forms by mail (from the initial information mailing) and from inside the questionnaires (obtained during the interview). When IRS 1040 and 1040A Forms were received from the IRS Service Centers, they were coded with case identifiers and entered in the master log, with the IRS coding batch number.

Westat made an additional check during the log-in procedure for the IRS returns. The receipt control clerk manually checked the list of dependents listed on the parents' return, and identified the student among them.

5.3.2 Coding

Coding manuals were prepared for each of the five data source documents used in the Pell Grant study for use in training the data preparation staff, serving as a complete and detailed reference for analysts, programmers, and data preparation staff and providing documentation for the Pell Grant study data files. Each coding manual consisted of an introduction to the study procedures and purposes, a review of general data preparation procedures to be followed, and coding and editing specifications for the five data sets.

Westat selected about 30 survey processing personnel for training as coding and editing staff for the Pell Grant study, including two group leaders, who were selected based on their skills and qualifications. The group leaders were assigned as assistants to the coding supervisors and as code verifiers. As much as possible, Westat chose experienced survey processing personnel as coders for this project to minimize the amount of training necessary on basic coding skills.

Coders were trained in groups of between 4 and 10. Each group was trained to code on one of the five study data sets. Between 1 and 8 hours of training time were required, depending on which data set was to be coded.

Questionnaires were precolumned before printing so that coding could be written directly on the questionnaires. FIRs, TARs and IRS 1040 and 1040A Forms could not be precolumned, so transcription sheets were designed for the coding.

Coders were assigned work by coding batch and were required to complete the coding of one batch before beginning work on another. Errors found during verification by the supervisor were first noted in a coding error log, then discussed with the coder committing them. If persistent errors were discovered, a coder was asked to review previous batches and correct them. Problems found during coding but not resolved in the coding specifications were documented and referred to a supervisor to be resolved. Particularly difficult cases were referred to a weekly meeting of senior project staff for resolution.

Occasional problems with illegible figures arose in the coding of photocopies of IRS tax forms. It was sometimes necessary in these situations to code illegible data elements as missing values. Some taxpayers do not completely fill out the 1040A Form when filing, exercising their option to have the IRS calculate their taxes. These 1040A Forms are blank below line 12. Westat coders were trained to fill in the missing items on these blank forms using a 1981 tax table.

The major coding problem for the Student and Parent Questionnaires was the large number of questions in each questionnaire which were open-ended. It was not possible prior to the beginning of coding to devise lists of all the possible responses to these items. Since this problem was expected, a controlled system of dealing with it

was implemented at the beginning of coding: responses which were not on the predetermined list of codes in the coding specifications were documented and referred to the supervisors who constructed codes for the new items. New codes were distributed each morning on a Coding Change Sheet. Coders were responsible for keeping their manuals up to date and were required to record each issue of the coding changes in a log.

5.3.3 Data Retrieval

Westat trained its coders to edit the data collection instrument during the coding. Editing involved checking for readability, sensibility, and following of skip patterns. (Editing was much more important in the coding of the questionnaires than in the coding of the secondary data sources.) A general rule was established that all primary verification questions in the questionnaire must have codable responses. When coders found erroneous skips, illegible answers, or illogical responses in any of the verification questions, they documented the problem and referred the case to a supervisor for data retrieval.

Experienced, specially trained telephone interviewers retrieved doubtful or missing data. Case problems were described on a Data Retrieval Request Form which also served as a record of calls for the interviewer. Data retrieval was attempted on 367 Parent Questionnaires and 407 Student Questionnaires. In addition to data retrieval due to problems found in the coding edit, data retrieval requests were also generated during machine editing.

5.3.4 Key Entry and Machine Editing

Coded documents and questionnaires which had been verified were taken to the Westat data entry staff in groups (called "keying batches") of approximately 100 documents. Coded documents were keyed into an in-house disk storage system, and then key verified from the disk. After keying and key verifying, the data were transmitted to Westat's VAX computer where they were stored on tapes to await machine editing.

All data sets except those for the Tax Assessor Records and Financial Institution Records were machine edited with special purpose COBOL programs, written to check for out-of-range codes, incorrect skip patterns, and inconsistent response patterns. The data sets were grouped by keying batch for editing.

Machine edit staff were trained coders, the majority of whom had earlier coded and edited the questionnaires as they were sent from the field. The coding supervisor, who had had extensive machine editing experience, also supervised the machine editing of the questionnaire files.

Machine edit clerks wrote file updating instructions on transcription sheets, supervisors checked these instructions and sent them to the data entry office for keying and transmittal to the computer center. Updates were made to the files by a special purpose COBOL update program. After each updating was complete, another editing cycle was run to verify that corrections had been made and to check for new errors. The update-edit cycle was repeated until each batch of data in the data set was clean.

Frequency distributions were run on all variables in each data set after the machine edit process was complete. The frequency distributions were carefully checked to uncover any inconsistencies and errors not found in the machine edit process. Any identified problems were corrected and new frequency distributions were run to assure accuracy.

6
DATA ANALYSIS

In the last chapter we discussed the processing of the data and the merger of all recipient-based data into a single file. This prepared the data for analysis, which itself was a complex process. Included in this chapter are descriptions of best value selection, detailed research questions, and table production and statistical analysis.

6.1 BEST VALUE SELECTION

The merged recipient data base is a compilation of cleaned and edited data from seven different data sources: Student Record Abstracts (SRA), Student Questionnaires (SQ), Parent Questionnaires (PQ), IRS Tax Returns (IRS), Financial Institution Records (FIR), Tax Assessor Records (TAR), and Computed Applicant Records (CAR) of the Student Aid Report (SAR). This merged data base, however, includes possibly different responses to identical items. For any one recipient responses to questions regarding Adjusted Gross Income, for example, might be contained in the SRA, PQ, SQ, or IRS. Thus, until further compilation of this data was performed, this raw data file was of limited use in analysis.

The basic analytic task of the study was to determine the difference between the "best" value for a given item and the value that appeared on the application. Any discrepancy for a student or parent item might have affected the Student Aid Index and thus the amount of the award. Discrepancies in application item values that had payment consequences contributed to program-wide estimates of error. But, to determine the existence of a discrepancy required the selection of one value from all of the possible sources as the best value. This was accomplished according to a detailed process of best value selection.

6.1.1 Strength Of Data Sources and Documentation

The process of best value selection was a combination of objective and subjective judgements about the relative strength of each possible source for a given application item. These were based on the views of Advanced Technology's senior staff, which included acknowledged experts in student financial aid and experienced analysts, with the corroboration of OSFA project staff and an outside consultant. The process was highly detailed and individually tailored for each application item. Decisions for best value selection were based jointly on the source of the data and the supporting documentation provided. Across data sources, there was a general hierarchy of strength which was followed. An external, reliable data source (IRS, FIR, TAR) was always considered first. If there was no such source present and documentation from other sources was of similar strength, the order followed was PQ or SQ (as applicable; for dependency status items where both PQ and SQ might contribute, the PQ preceeded the SQ), SRA, and, if no other source was present, the application value from the CAR.

For supporting documentation, a separate hierarchy was established for each application item. In general, values documented by copies of directly relevant external records were first, such as a copy of a recent real estate appraisal for house value. These were followed by values with documentation from other external sources deemed to be less directly relevant or timely, such as a copy of a document showing property insurance held on a home, or a two year old bill-of-sale on a home. The next level of documentation was a statement from a knowledgeable professional (such as an accountant), followed by personal records from the respondent or a notarized statement from the respondent. Some documentation was considered unacceptable and was never used, since it was considered less reliable than the application. An example would be data retrieved by telephone or an irrelevant item from a tax return.

When the general hierarchies for source and documentation are combined it is easy to see the complexity of the best value selection process. This is illustrated in Figures 6-1, 6-2, and 6-3 which show the priorities used for Adjusted Gross Income (for independent students), Home Value (for parents' of dependent students), and Supported by Parents, 1981. The first two figures provide different examples of items for which documentation is often present, the third figure illustrates the subtleties involved in

CONDITION	PRIORITY	SOURCE	QUESTION NUMBER	CODE	LABEL
If *FTAX = filed or missing; If 1. *MARRY = not married and *FSTAT = R only or 2. *MARRY = married and *FSTAT = joint or *FSTAT = married separate and both rec'd	1.	IRS/ SQ	L.31 - SQ39a	-	1040 minus CW-S
			L.10 - SQ39a	-	1040A minus CW-S
	2.	SQ	38 - 39a	01	1040 Cert. minus CW-S
				02	1040A Cert. minus CW-S
	3.	SRA	19	01	1040 Cert.
				02	1040A Cert.
				10	IRS Transcript
	4.	SQ	38 - 39a	03	1040 minus CW-S
				04	1040A minus CW-S
	5.	SRA	19	03	1040
				04	1040A
				13	Puerto Rican tax return
If 1. *MARRY = not married and *FSTAT = R only or 2. *MARRY = married and *FSTAT = joint or *FSTAT = married separate and both rec'd	6.	IRS	L.31	-	1040
			L.10	-	1040A

6-3

FIGURE 6-1

BEST VALUE SELECTION PRIORITIES:
ADJUSTED GROSS INCOME FOR INDEPENDENT STUDENTS

CONDITION	PRIORITY	SOURCE	QUESTION NUMBER	CODE	LABEL
If *FTAX = filed or missing	7.	SQ	38 - 39a	12	State Tax Return-Cert. minus CW-S
				19	City Tax Return-Cert. minus CW-S
				18	State Tax Return minus CW-S
	8.	SRA	19	08	State Tax Return
	9.	SQ	38 - 39a	17	IRS receipt/Treasury Dept. Statement minus CW-S
	10.	SRA	19	11	IRS letter
				14	1040 (no line number indicated)
				15	Separated student's portion of 1040
	11.	SQ	38 - 39a	10	W-2 minus CW-S
	12.	SRA	19	09	W-2
	13.	SQ	38 - 39a	20	1040X minus CW-S
	14.	SRA	19	12	1040X
				06	1040A worksheet
	15.	SQ	38 - 39a	11	Pay stub minus CW-S
				14	Statement from professional minus CW-S
				15	Statement from social agency minus CW-S
	16.	SRA	19	05	Statement from social agency
				07	Notarized statement
	17.	SQ	38 - 39a	16	Personal records
	18.	SAR	22		

FIGURE 6-1 (Continued)

BEST VALUE SELECTION PRIORITIES:
ADJUSTED GROSS INCOME FOR INDEPENDENT STUDENTS

CONDITION	PRIORITY	SOURCE	QUESTION NUMBER	CODE	LABEL
UNACCEPTABLE		SQ	38	85	Tax form listed
				86	Not applicable due to phone retrieval
				94	Partial documentation
				97	No documentation
				98	Don't know
				99	Not ascertained
		SRA	19	99	Not ascertained

FIGURE 6-1 (Continued)

BEST VALUE SELECTION PRIORITIES:
ADJUSTED GROSS INCOME FOR INDEPENDENT STUDENTS

CONDITION	PRIORITY	SOURCE	QUESTION NUMBER	CODE	LABEL
If value is greater than *HOMEP or *HOMEP missing	1.	TAR	QM Value		Tax assessor's record of market value
	2.	PQ	63	01	Appraisal
				02	Statement from local real estate office
	3.	SRA	29	01	Appraisal
				02	Statement from local real estate office
	4.	PQ	63	03	Tax assessment forms
	5.	SRA	29	03	Tax assessment forms
	6.	PQ	63	04	Property insurance
If PQ62b ≤ 03	7.	PQ	63	10	Mortgage statement
				11	Purchase contract
				12	Registration Certificate for Mobile home
				15	HUD form
				16	Loan application
				17	Deed
				18	Cancelled check for full amount
				19	Sales agreement
				20	Copy of statement of loans on property
				21	Closing statement
				22	Respondent's personal records
				23	Statement from social agency
				24	Title transfer
If greater than SAR or PQ62b ≤ 03	8.	*HOMEP			Purchase price of home
If *NHOME value = 0	9.	*NHOME			Proof that no home owned

FIGURE 6-2

BEST VALUE SELECTION PRIORITIES:
HOME VALUE FOR PARENT(S) OF DEPENDENT STUDENTS

HOME VALUE FOR PARENT(S) OF DEPENDENT STUDENTS

CONDITION	PRIORITY	SOURCE	QUESTION NUMBER	CODE	LABEL
	10.	SAR	37		
UNACCEPTABLE		PQ	63	13	Estimate based on document shown to interviewer
				14	Tax form
				85	Tax form listed
				86	Not applicable due to phone retrieval
				94	Partial documentation
				97	No documentation
				98	Don't know
				99	Not ascertained
		SRA	29	99	Not ascertained

FIGURE 6-2 (Continued)

BEST VALUE SELECTION PRIORITIES:
HOME VALUE FOR PARENT(S) OF DEPENDENT STUDENTS

CONDITION	PRIORITY	SOURCE	QUESTION NUMBER	CODE	LABEL
If values on all three sources agree	1.	PQ and	27	-	No documentation requested
		SQ and	30	-	No documentation requested
		SRA	14	01-08	See priorities 5 and 7 below
If values on PQ and one other source agree	2.	PQ and	27	-	No documentation requested
		SQ and	30	-	No documentation requested
		SRA	14	01-08	See priorities 5 and 7 below
	3.	PQ	27	-	No documentation requested
If values on both agree	4.	SQ and	30	-	No documentation requested
		SRA	14	01-08	See priorities 5 and 7 below
	5.	SRA	14	01	Deeds
				02	Title transfers
				03	Cancelled checks
				04	Notarized statement from parents
				05	Notarized statement from student
				06	Letter from social service agency
	6.	SQ	30	-	No documentation requested
	7.	SRA	14	07	Statement from parent
				08	Statement from student
	8.	SAR	14a		
UNACCEPTABLE		SRA	14	99	Not ascertained

FIGURE 6-3

BEST VALUE SELECTION PRIORITIES:
SUPPORT BY PARENTS 1981 FOR ALL STUDENTS

an item for which little documentation is typically available but a best value must be determined. Note that in the absence of documentation in Figure 6-3, agreement from multiple undocumented sources is considered strongest.

Best value selection must follow a closely prescribed order since the best values for some items were needed to determine the best value of other items. Consider, for example, that one must know the best dependency status, marital status, and tax filing status in order to select the best Adjusted Gross Income. (If the best dependency status is dependent, and the parents are separated but filed a joint tax return for 1981, that tax return cannot be used for AGI without additional documentation regarding the proportion of income received by the supporting parent.)

A list of all items for which best value priorities were developed appears in Figure 6-4. Since separate priorities were sometimes required by dependency status, these are listed by dependency status.

6.1.2 Premises in Best Value Selection

The selection of the best value from among several competing sources is limited by certain inherent restrictions in the nature of the Pell Quality Control study. Our aim is to verify the application values against the best sources at our disposal. Thus, our best value, even if it is from the highest priority source, is only best relative to others available. We cannot investigate the veracity of the sources we use. Thus, an IRS-certified copy of a tax return is our best source for several items, based on the assumption that the tax return is accurate. If that tax return was itself inaccurate, for whatever reason, our best value would be inaccurate.

A basic premise of best value selection is that application values be changed only when there is evidence of a value from a better source. Thus, in the absence of documentation we presume that the application value is correct. This presumption may occasionally lead to the rejection of a more accurate but undocumented value and the underestimation of absolute error. However, it is based on the simple and logical assertion that accurate recall of undocumented values was better at the time of application, when the item in question was current, than it would be more than one year later.

ALL STUDENTS

Determining Student's Dependency Status
Expected Social Security Educational Benefits
Expected VA Educational Benefits
Marital Status
Support By Parents 1981
Support By Parents 1982
Claimed By Parents 1981
Claimed By Parents 1982
Lived With Parents 1981
Lived With Parents 1982
Child Support
Other Weifare
Non-Educational VA Benefits
Untaxed Unemployment Compensation
Interest on Tax Free Bonds
Untaxed Pensions and Capital Gains
Untaxed Housing Allowance
Earnings Not Reported on Tax Return
Interest/Dividend Exclusion
Any Other Income
IRS Filing Status
Applicant's Expected Taxable Summer Income
Applicant's Expected Taxable School Year Income
Spouse's Expected Taxable Summer Income
Spouse's Expected Taxable School Year Income

INDEPENDENT STUDENTS ONLY

Filed/Did Not File Tax Return
Adjusted Gross Income
U.S. Taxes Paid
State/Local Taxes Paid
Applicant's Income
Spouse's Income
AFDC
Other Income/Benefits Total
Number of Exemptions
Household Size
Number in Postsecondary Education
Medical/Dental Expenses
Elementary/Secondary Tuition
Cash/Savings/Checking
Purchase Price of Home
No Home Owned
Home Value
Home Debt
Investment Value
Investment Debt
Farm Value
Business Value
Business/Farm Value
Business Debt
Farm Debt
Business/Farm Debt
Itemized Deductions

FIGURE 6-4

LIST OF ITEMS WITH BEST VALUE SELECTION PRIORITIES
6-10

DEPENDENT STUDENTS ONLY

Filed/Did Not File Tax Return
Adjusted Gross Income
U.S. Taxes Paid
State/Local Taxes Paid
AFDC
Other Income/Benefits Total
Cash/Savings/Checking
Purchase Price of Home
No Home Owned
Home Value
Home Debt
Real Estate/Investment Value
Real Estate/Investment Debt
Farm Value
Business Value
Business/Farm Value
Farm Debt
Business Debt
Business/Farm Debt
Dependent Student's (and Spouse's) Net Income
Dependent Student's (and Spouse's) Net Assets
Applicant's Income
Spouse's Income

PARENT(S) OF DEPENDENT STUDENTS

Filed/Did Not File Tax Return
IRS Filing Status
Adjusted Gross Income
U.S. Taxes Paid
State/Local Taxes Paid
Father's Income
Mother's Income
AFDC
Child Support
Other Welfare
Non-Educational VA Benefits
Untaxed Unemployment Compensation
Interest on Tax Free Bonds
Untaxed Pensions and Capital Gains
Untaxed Housing Allowance
Earnings Not Reported on Tax Return
Interest/Dividend Exclusion
Any Other Income
Other Income/Benefits Total
Number of Exemptions
Household Size
Number in Postsecondary Education
Medical/Dental Expenses
Elementary/Secondary Tuition
Cash/Savings/Checking
Purchase Price of Home
No Home Owned
Home Value
Home Debt
Real Estate/Investment Value
Real Estate/Investment Debt
Farm Value
Business Value
Business/Farm Value
Business Debt
Farm Debt
Business/Farm Debt
Marital Status
Parent's Social Security Benefits
Itemized Deductions

FIGURE 6-4 (Continued)

LIST OF ITEMS WITH BEST VALUE SELECTION PRIORITIES

There were a few exceptions to this premise of not accepting undocumented values as best values. These were applied systematically to items which were composite totals on the application for which the separate parts were not reported to the central processor. The most important of these was Other Nontaxable Income. Individual contributing elements to Other Nontaxable Income, such as child support or untaxed unemployment compensation, were recorded on the application worksheet, but only the total was reported on the application itself. Thus, our efforts to select the best value based on documentation of the separate parts could not proceed unless a provision was made for a default value to be used when no documentation was available. Recall that individually reported items used the application value as a default value. When a person reported during the interview the receipt of $2500 in undocumented child support we could not use a value of zero without grossly underreporting income. Nor was there any alternative value to be picked up from the application, since the application value included nontaxable income from all sources. Thus, the only reasonable approach was to accept the undocumented report.

Some other items also required the acceptance of undocumented values because the separate contributing parts were not reported on the application. These were the contributing parts to Dependent Student's (and Spouse's) Net Income and Dependent Student's (and Spouse's) Net Assets. In addition, to avoid large overestimations of net worth, we accepted undocumented values of Home Debt, Real Estate/Investment Debt, Business Debt, or Farm Debt when corresponding documented values for Home Value, Real Estate/Investment Value, Business Value, or Farm Value were present.

6.2 DETAILED RESEARCH QUESTIONS

Given the amount of data collected it was necessary to determine in advance those areas that would be the focus of data analysis. While all analysis of error followed the general equations set forth in the Appendix of Volume 1, the number of specific error items and comparisons between items was so large that potential research questions were delineated in advance and submitted to the project officer before analysis began. A total of 145 research questions were prepared, divided by topics likely to be included as chapters in Volume 1. These specific research questions presented rates and amounts of error (e.g., What is the marginal impact of application item error on net student error?), bivariate interactions or crosstabulations of error with other variables (Does institutional error differ by type and control of

institution?), and multivariate analyses (What are the contributing factors of student error?). For each specific question the measures that would be used were delineated; for bivariate and multivariate analyses the independent and dependent variables were listed separately.

These research questions served as a guideline for the subsequent analysis and the production of tables for the findings. Based on the initial results of the early analyses, specific research questions were revised, expanded, or, in some cases, dropped. Together these research questions enabled us to pursue data analysis in an organized fashion. Each of the tables presented in Volume I represents the findings of one or more specific research questions.

6.3 STATISTICAL ANALYSIS AND TABLE PRODUCTION

Before statistical analysis of the data began, a master analysis file was created. This file included the 3175 cases that were considered complete for analytical purposes and data for 311 institutions.* It represents a merger of the final recipient-based data file, which included best values and selected individual items from various data sources, institutional data for individual recipients and the institution as a whole, and nonresponse adjustment weights for each recipient. (For a discussion of complete cases and nonresponse adjustment see Chapter 7 of Volume 1.)

All analysis was conducted using SAS, a comprehensive package for statistical analysis. SAS had also been used for the best value selection and the creation of all files used in analysis. The features of SAS, especially some new procedures included in the recently released 1982 version, facilitated the accurate production of tables.

6.4 CONTROL GROUPS

One potential threat to the validity of the study was experimental bias--the impact of selection for participation in the study on the findings. In order to assess

*While 317 institutions participated, six of them were branch campuses which maintained no separate financial aid offices. Therefore, the financial aid office for the central campus, which in all cases was included in our sample, served for the branch as well.

the impact of experimental bias on the study two control groups were established. One, the Institutional Control Group (ICG), consisted of students in the sampled institutions who were not selected for the study. Student aid files for these students, whose identity was not known to the institutions, were reviewed during the spring data collection. These ICG participants were not interviewed, so comparisons could be made only on items related to institutional error.

A second control group consisted of students whose records were randomly selected from the Computed Applicant Records of the Student Aid Report. Students selected for this control group, which involved computer tape data review only, were from both sampled and nonsampled institutions.

The use of these two control groups to assess experimental bias is detailed in Chapter 7 of Volume 1, along with the relevant findings. Data for experimental bias assessment was analyzed using both SAS and SPSS, another statistical package.

7

SURVEY RESPONSE RATES

As in every voluntary survey, it was not possible to gain the cooperation of every institution to permit the drawing of a sample of its students or of every student and parent sampled at cooperating institutions. Nevertheless, we did achieve high survey response rates at both the institutional and individual levels.

7.1 INSTITUTIONAL SURVEY RESPONSE RATES

The institutional sample initially drawn included 322 schools and colleges. For various reasons, five could not or would not cooperate with the survey. Therefore, the institutional response rate was 98.4 percent.

The first occasion for institutional refusal occurred during the telephone calling to schedule the fall round of visits. Telephone schedulers who encountered refusals referred them to their Westat supervisor, who could refer them to Advanced Technology's institutional data collection manager, who was at the Westat telephone center to make rescheduling decisions. The data collection manager converted several refusals, including some based on an incorrect interpretation of the Buckley Amendment. Some refusals were referred to the project officer for resolution. The five ultimate refusals were made for the following reasons:

- At one public four-year institution, a fire in the administration building had destroyed the financial aid records.
- At two proprietary institutions, all financial aid matters were handled by a consultant who was abusively hostile, as he had been when two of his clients were included in the Stage One sample.
- At one proprietary institution, the financial aid director and the director of the school were parties to litigation under trial and could not be available for the fall visit. They also refused to cooperate with the spring visits.

- At one proprietary institution, a new financial aid director was already in the midst of a set of audits by accrediting organizations and other government agencies when we first contacted her. She also refused cooperation when recontacted in the spring.

7.2 STUDENT AND PARENT SURVEY RESPONSE

The number of completed cases was determined not only by the proportion of students and parents eligible for the study who cooperated by granting an interview, but also by what proportion of the initial sample we eventually determined not to have been part of the population being studied. We present the overall response rate first and then explain the problem of sample loss.

7.2.1 Survey Response Rate

The response rates presented in this section represent "final" disposition categories determined by the coders when the questionnaires were edited and coded, after receipt at Westat's home office. In some cases, the disposition of an apparently completed questionnaire was changed as a result of additional work done at the home office after the questionnaire had been sent in as a completed instrument. For example, if a completed student questionnaire showed evidence that the student filed a "Special Condition Application" in applying for a Pell Grant, the financial aid administrator (FAA) at the student's institution was telephoned. If the FAA indicated that the student did file a Special Condition Application, the status of the questionnaire was changed from "complete" to "sample loss," because special condition filers should not have been in the sample. If the parent of that student also completed an interview, the parent's questionnaire was also changed from "complete" to "sample loss."

Calculating the response rates from the clean data file represents a very conservative approach to specifying survey response statistics. Since the response rates cited in this section summarize the status of each questionnaire as it was edited by the home office coder, they slightly understate the response rates when compared to the operational response rates calculated during the field period and reported weekly to the Department of Education.

The dispositions of all the sampled students and parents are shown in Table 7-1. The percentages in the table represent the proportion of completed interviews of those eligible, sampled students and parents where an interview was possible. Ineligible sample members, principally students (and therefore their parents) who never actually received a Pell Grant or who filed a Special Condition Application, and sampled individuals who were dead or out of the country were deleted from the initial sample used as the denominator in the percentages. A separate analysis of this sample loss category has been included in the report on findings as a part of the analysis of non-response.

To satisfy the analytic needs of the Pell Grant QC study, both a parent and a student questionnaire must be completed for dependent student grants, while a student questionnaire alone is sufficient to verify an independent student grant. The last row of Table 7-1 reflects a combined response rate for pairs of dependent students and parents and for single independent students.

7.2.2 Sample Loss

The initial sample of 4,109 students was drawn early in the school year from two groups of students: those who were already receiving Pell Grants, and those who had been certified as eligible pending successful validation by the financial aid offices at their schools. We made efforts to identify and remove from the sampling frame students who did not receive a grant or who had used a Special Condition Application when they applied. It was, however, often impossible to identify these ineligible sampled students, particularly among the group of students pending validation. Table 7-2 shows the number of student and parent cases and dispositions that make up the sample loss category. Because it includes both students and parents, the survey sample loss count is greater than the case sample loss count of 422 reported in Chapter 7 of Volume 1.

The sampled-in-error category represents students who were sampled in the fall as recipients or recipients pending validation, but who never actually received a grant, or who never actually attended the institution where they were sampled. If students were found to be in any of the sample loss categories, their parents were also included in that category, and both were removed from the data file.

TABLE 7-1

PRELIMINARY STUDENT/PARENT SURVEY RESPONSE RATES

Grant Type and Form of Questionnaire	Number in Sample	Number Ineligible**	Net Sample	Number Completed	Percent Completed
Dependent Student Grants:					
Student	2,388	198	2,190	2,136	97.5
Parent	2,388	211	2,177	2,094	96.2
Paired Student and Parent*			4,332	4,140	95.6
Independent Student Grants:					
Student	1,690	145	1,545	1,440	93.2
Parent	1,690	355	1,335	967	72.4
Paired Student and Parent*			2,638	1,610	61.0
Dependency Status Not Determined*	62	59	3	0	
Dependent and Independent Grants:					
Paired Student and Parent	8,218		6,970	5,750	82.5
Paired Student and Parent for dependent student grants and student only for independent student grants	6,466		5,879	5,580	95.0

*Count of student and parent questionnaire (not a count of cases)
**Sample loss

Although the sample loss represents a reduction in the overall sample size for the survey, it is well within acceptable limits. Since the number of completed interviews exceeded the target, the precision of the sample estimates is not seriously affected by the reduction.

TABLE 7-2

DISPOSITION OF STUDENT AND PARENT SURVEY SAMPLE LOSS CASES

Disposition Category	Number in Category	Percent of Sample Loss
Sampled in Error	465	48.1%
Out of Country	161	16.6%
Deceased	121	12.5%
Special Condition Filer	221	22.8%
Total	**968**	**100.0%**

7.2.3 Response Rate for Analytical Purposes

The response rates presented in this chapter reflect the success of the efforts to interview students and their parents. Data analysis, however, requires not only that an interview be conducted but that all sections of the interview be complete and all other records be available. Thus, some cases that were considered complete for survey purposes have been recoded as incomplete for analytical purposes, reducing the survey response rate of almost 95 percent to a useful response rate of over 86 percent. This reduction was primarily accounted for by two reasons.

- Parents of independent students were requested to complete only part of the interview, with income and asset information omitted. Because their participation was completely voluntary and these data would not be needed in most cases, this information was excluded to facilitate cooperation and protect their privacy. If, however, the student was later deemed to be dependent, the omitted parent information became essential for a case to be considered complete for analysis. In some instances, it was possible to obtain this additional information from parents; in most (116 cases) it was not. This reduced the functional response rate.

- For some recipients we could not obtain their Computed Applicant Record from the Student Aid Report, without which error could not be determined. The reason for this was a failure to match the Social Security number (SSN) on our records (which were verified during the interview) with either the original SSN or the current SSN on the SAR file. This occurred in 72 cases. Some of these failures to match were due to a damaged tape (the tape from the central processor was received with unreadable data on about one half of one percent of the blocks, which would be 16 cases if evenly distributed), the remainder did not match for unknown reasons.

Additional details on the functional response rate are given in Chapter 7 of Volume 1.

7.3 VARIANCE ESTIMATES

A critical part of the findings in Volume 1 are program-wide estimates of error in the population of Pell recipients. These estimates, of course, may vary from the actual population figures to the extent that the sample differs from the population. A series of 42 tables presents selected estimates, standard errors of the estimates, and coefficients of variation.

7.3.1 Variance Estimation by the Method of Balanced Repeated Replication

The sampling error of an estimate, based on any sample design using any estimation procedure, no matter how complex, may be estimated by the method of replications. Theoretically, this method is equivalent to the idea that the sample selection, collection of data, and estimation procedures are independently carried through (replicated) several times. In practice, random 50 percent subsets of the survey results are selected and estimates formed from each. The dispersion of the resulting estimates can be used to measure the variance of the full sample.

The method of replications has special advantages in reducing the complexity of variance computations. Another benefit is that it may be applied to compute sampling errors for higher-order statistics without the need for new variance expressions.

The method consists of three steps:

- Assemble data for the sample units that make up each of the replicates. This is equivalent to making a copy of the sample data for the units in each of the subsamples of the full sample.

- Perform the estimation procedure on each of the replicates. The same estimation procedure, prepared for the full sample, is applied separately to each of the replicates.

- Calculate the dispersion of the resulting estimates among the replicates to estimate the variance of the full sample; a relatively simple computation formula is used that does not depend on the form of the estimate for which the variance is to be approximated.

Definition of the Replicates. Each of the half-sample replicates prepared for variance estimation must satisfy two criteria:

- The replicate must comprise a sample approximately one-half the size of the full sample.

- The selection of the half-sample must observe the same sampling principles as the full sample.

For the replicates defined for the Pell Grant design, these criteria are satisfied by selecting half-samples of the units designated in the first stage of sampling in the full sample. For nonself-representing institutions, a replicate comprises all students selected in half of the clusters; in self-representing institutitons a replicate comprises half of the students selected at the school. Half of the Alternative Disbursement System (ADS) students were selected in each half-sample.

Each of the 34 certainty institutions is treated as a stratum. Half of the students in each institution were assigned half-sample code 1 and half were assigned half-sample code 2. All ADS students were placed in stratum 35 with half assigned a code of 1 and half assigned a code of 2.

Students in noncertainty institutions were assigned to a stratum and half-sample based on the geographic cluster of the institution. Clusters were paired (in the order they appeared in the sampling frame) to form 36 strata. All students in a cluster were assigned the same half-sample code (1 or 2).

Half-samples of the full sample were defined by randomly selecting one or the other half-sample code from each of the 71 pairs; the number of different half-sample replicates possible by this method would equal 2 raised to the 71st power (abouth 2,361 billion). McCarthy has shown that the variance can be estimated with equivalent

reliability from only a small number of orthogonal replicates.* For the Pell Grant program, the number of orthogonal replicates needed is the smallest multiple of four equal to or greater than the number of pairs. With 71 pairs, the number of replicates needed is 72.

Variance Calculations. For every estimate X calculated from the total sample, let x_i be the estimate calculated from the ith half-sample, i = 1, 2, ..., 72. The variance estimate for X is then

$$\text{Var}(X) = \frac{1}{72} \sum_{i=1}^{72} (x_i - X)^2.$$

This formula has been used to compute the standard errors (square root of the variance) for each statistic presented in Tables 7-3 through 7-44.

7.3.2 Estimated Sampling Errors

In this section we present estimated sampling errors developed using the methods described in the previous section. For each statistic we present the estimate itself, the standard error of the estimate, and the coefficient of variation (standard error of the estimate divided by the estimate).

Table 7-3 presents the standard errors for program-wide dollar estimates of error. The coefficients of variation for student and case error are always less than 12 percent, and the coefficients are often as low as 6 percent. Estimated standard errors for number of cases with payment or eligibility errors are presented in Table 7-4. The standard errors for estimated numbers of payment errors are at or below 8 percent. Occurrence of eligibility errors, except no Financial Aid Transcript, was quite rare. Therefore, the coefficients of variation will appear to be high. However, the standard errors are low enough so that reasonably tight confidence intervals exist.

---

*U.S. Department of Health, Education and Welfare, National Center for Health Statistics. Replication: An Approach to the Analysis of Data from Complex Surveys by Philip J. McCarthy, Series 2, No. 14, Washington, D.C., Government Printing Office.

TABLE 7-3

ESTIMATED SAMPLING ERRORS:
SELECTED PROGRAM-WIDE DOLLAR ESTIMATES

Statistic	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Overaward			
Institution	206.36	26.68	0.13
Student	265.25	13.95	0.05
Case	452.43	29.00	0.06
Underaward			
Institution	104.25	18.79	0.18
Student	54.33	5.67	0.10
Case	136.20	15.91	0.12
Net Error			
Institution	102.11	34.21	0.34
Student	210.92	15.52	0.07
Case	316.22	35.99	0.11
Absolute Error			
Institution	310.62	30.98	0.10
Student	319.58	14.58	0.05
Case	588.63	29.88	0.05

TABLE 7-4

ESTIMATED SAMPLING ERRORS:
SELECTED NUMBER OF RECIPIENTS WITH ERROR

Statistic	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Institutional Overaward	388.05	28.23	0.07
Institutional Underaward	437.20	33.17	0.08
Student Overaward	758.07	30.08	0.04
Student Underaward	218.81	13.47	0.06
Case Overaward	1,022.72	31.90	0.03
Case Underaward	525.29	24.71	0.05
ELIGIBILITY ERRORS			
Unsatisfactory Academic Progress	9.22	3.68	0.40
Less Than Half Time	2.98	1.48	0.50
Insufficient Program Length	0.76	0.73	0.95
Nondegree Program	11.89	1.72	0.14
No Statement of Educational Purpose	9.38	3.45	0.37
No Financial Aid Transcript	82.21	12.49	0.15
Loan Default	0.70	0.66	0.96
Invalid SAR	9.92	2.92	0.29
Not Parent School	5.10	2.16	0.42

Standard errors for mean payment errors are presented in Table 7-5. These standard errors are generally less than $25; thus, 95 percent confidence intervals would generally have a width of less than $100.

The remaining tables provide sampling errors for the various estimates for the following categories:

Type and Control of Institution
Grant Type (Dependency Status)
Validation Status

Tables 7-6 to 7-17 present estimates for the various program-wide estimates of error. Tables 7-18 to 7-23 show estimates for the number of recipients with errors. Tables 7-24 to 7-35 contain sampling errors for mean payment errors; and, finally, Tables 7-36 to 7-44 show the estimated number of recipients with eligibility errors.

TABLE 7-5

ESTIMATED SAMPLING ERRORS:
SELECTED MEAN DOLLAR ESTIMATES

Statistic	Estimate ($)	Standard Error ($)	Coefficient of Variation
Mean Overaward			
Institution	532	45	0.08
Student	350	11	0.03
Case	442	23	0.05
Mean Underaward			
Institution	238	33	0.14
Student	248	15	0.06
Case	259	22	0.09
Mean Net Error			
Institution	40	14	0.34
Student	83	6	0.07
Case	125	14	0.11
Mean Absolute Error			
Institution	123	12	0.10
Student	126	6	0.05
Case	233	12	0.05

TABLE 7-6

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
INSTITUTIONAL OVERAWARD

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	90.75	21.60	0.24
Private 4 Year	49.98	11.56	0.23
Public 2 Year	25.15	3.82	0.15
Private 2 Year	2.79	1.80	0.65
Proprietary 2 Year	20.69	15.22	0.74
Proprietary Less Than 2 Year	17.00	2.63	0.15
Grant Type			
Independent	123.27	21.81	0.18
Dependent	83.09	11.16	0.13
Validation Status			
Non-Validated	86.42	12.05	0.14
Validated	119.94	18.78	0.16
All Categories Combined	206.36	26.68	0.13

TABLE 7-7

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
STUDENT OVERAWARD

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	130.16	13.29	0.10
Private 4 Year	63.24	9.03	0.14
Public 2 Year	41.92	5.76	0.14
Private 2 Year	6.06	2.29	0.38
Proprietary 2 Year	12.18	5.53	0.45
Proprietary Less Than 2 Year	10.76	3.39	0.31
Grant Type			
Independent	104.09	10.86	0.10
Dependent	161.16	8.24	0.05
Validation Status			
Non-Validated	91.43	6.78	0.07
Validated	173.82	13.35	0.08
All Categories Combined	265.25	13.95	0.05

TABLE 7-8

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASE OVERAWARD

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	217.61	26.84	0.12
Private 4 Year	107.87	16.82	0.16
Public 2 Year	61.73	7.06	0.11
Private 2 Year	8.51	3.00	0.35
Proprietary 2 Year	30.51	17.42	0.57
Proprietary Less Than 2 Year	26.20	4.30	0.16
Grant Type			
Independent	217.76	23.26	0.11
Dependent	234.66	12.55	0.05
Validation Status			
Non-Validated	168.44	15.29	0.09
Validated	283.99	21.52	0.08
All Categories Combined	452.43	29.00	0.06

TABLE 7-9

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
INSTITUTIONAL UNDERAWARD

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	25.82	3.43	0.13
Private 4 Year	14.95	4.11	0.27
Public 2 Year	39.98	15.50	0.39
Private 2 Year	3.34	2.53	0.76
Proprietary 2 Year	7.70	3.15	0.41
Proprietary Less Than 2 Year	12.40	6.33	0.51
Grant Type			
Independent	48.81	12.19	0.25
Dependent	55.45	7.95	0.14
Validation Status			
Non-Validated	46.97	9.76	0.21
Validated	57.28	9.51	0.17
All Categories Combined	104.25	18.79	0.18

TABLE 7-10

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
STUDENT UNDERAWARD

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	25.81	4.54	0.18
Private 4 Year	15.62	2.97	0.19
Public 2 Year	8.01	2.28	0.28
Private 2 Year	1.87	1.06	0.56
Proprietary 2 Year	1.56	0.76	0.48
Proprietary Less Than 2 Year	1.30	0.67	0.52
Grant Type			
Independent	11.86	3.26	0.27
Dependent	42.48	4.63	0.11
Validation Status			
Non-Validated	17.52	3.28	0.19
Validated	36.82	3.70	0.10
All Categories Combined	54.33	5.67	0.10

TABLE 7-11

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASE UNDERAWARD

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	44.57	5.51	0.12
Private 4 Year	25.34	4.59	0.18
Public 2 Year	42.33	13.88	0.33
Private 2 Year	4.87	3.00	0.62
Proprietary 2 Year	6.89	2.52	0.36
Proprietary Less Than 2 Year	12.20	6.07	0.50
Grant Type			
Independent	50.74	9.79	0.19
Dependent	85.46	8.47	0.10
Validation Status			
Non-Validated	55.22	8.93	0.16
Validated	80.99	8.14	0.10
All Categories Combined	136.20	15.91	0.12

TABLE 7-12

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NET INSTITUTIONAL ERROR

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	64.94	21.42	0.33
Private 4 Year	35.03	13.24	0.38
Public 2 Year	-14.83	16.34	1.10
Private 2 Year	-0.55	2.96	5.36
Proprietary 2 Year	12.99	14.93	1.15
Proprietary Less Than 2 Year	4.54	7.23	1.59
Grant Type			
Independent	74.47	25.82	0.35
Dependent	27.64	14.74	0.53
Validation Status			
Non-Validated	39.45	16.44	0.42
Validated	62.66	21.73	0.35
All Categories Combined	102.11	34.21	0.34

TABLE 7-13

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
ABSOLUTE INSTITUTIONAL ERROR

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	116.57	22.32	0.19
Private 4 Year	64.93	11.21	0.17
Public 2 Year	65.13	15.59	0.24
Private 2 Year	6.13	3.26	0.53
Proprietary 2 Year	28.39	16.13	0.57
Proprietary Less Than 2 Year	29.46	6.46	0.22
Grant Type			
Independent	172.08	24.12	0.14
Dependent	138.54	12.57	0.09
Validation Status			
Non-Validated	133.39	14.51	0.11
Validated	177.23	20.36	0.11
All Categories Combined	310.62	30.98	0.10

TABLE 7-14

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NET STUDENT ERROR

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	104.35	13.40	0.13
Private 4 Year	47.62	8.70	0.18
Public 2 Year	33.91	5.41	0.16
Private 2 Year	4.18	1.87	0.45
Proprietary 2 Year	10.62	5.00	0.47
Proprietary Less Than 2 Year	9.46	3.18	0.34
Grant Type			
Independent	92.23	12.00	0.13
Dependent	118.68	9.01	0.08
Validation Status			
Non-Validated	73.92	7.97	0.11
Validated	137.00	13.93	0.10
All Categories Combined	210.92	15.52	0.07

TABLE 7-15

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
ABSOLUTE STUDENT ERROR

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	155.98	14.66	0.09
Private 4 Year	78.86	10.25	0.13
Public 2 Year	49.93	6.89	0.14
Private 2 Year	7.93	3.03	0.38
Proprietary 2 Year	13.75	6.11	0.44
Proprietary Less Than 2 Year	12.06	3.71	0.31
Grant Type			
Independent	115.94	10.63	0.09
Dependent	203.64	9.89	0.05
Validation Status			
Non-Validated	108.95	7.07	0.06
Validated	210.63	13.77	0.07
All Categories Combined	319.63	14.58	0.05

TABLE 7-16

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NET CASE ERROR

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	173.03	25.11	0.15
Private 4 Year	82.54	17.72	0.21
Public 2 Year	19.41	15.48	0.80
Private 2 Year	3.63	2.73	0.75
Proprietary 2 Year	23.61	16.94	0.72
Proprietary Less Than 2 Year	14.00	7.24	0.52
Grant Type			
Independent	167.02	25.19	0.15
Dependent	149.21	16.25	0.11
Validation Status			
Non-Validated	113.22	19.26	0.17
Validated	203.00	24.54	0.12
All Categories Combined	316.22	35.99	0.11

TABLE 7-17

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
ABSOLUTE CASE ERROR

Category	Estimate ($ Millions)	Standard Error ($ Millions)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	262.18	29.52	0.11
Private 4 Year	133.21	17.15	0.13
Public 2 Year	104.04	15.66	0.15
Private 2 Year	13.38	5.35	0.40
Proprietary 2 Year	37.40	18.24	0.49
Proprietary Less Than 2 Year	38.40	7.64	0.20
Grant Type			
Independent	268.51	25.29	0.09
Dependent	320.12	13.96	0.04
Validation Status			
Non-Validated	223.65	15.99	0.07
Validated	364.97	21.37	0.06
All Categories Combined	588.63	29.88	0.05

TABLE 7-18

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH INSTITUTIONAL OVERAWARD

Category	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	170.97	25.09	0.15
Private 4 Year	60.49	14.49	0.24
Public 2 Year	103.69	14.10	0.14
Private 2 Year	4.69	2.05	0.44
Proprietary 2 Year	23.25	15.77	0.68
Proprietary Less Than 2 Year	24.97	4.64	0.19
Grant Type			
Independent	207.99	21.77	0.10
Dependent	180.06	14.81	0.08
Validation Status			
Non-Validated	161.04	14.96	0.09
Validated	227.01	19.76	0.09
All Categories Combined	388.05	28.23	0.07

TABLE 7-19

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH INSTITUTIONAL UNDERAWARD

Category	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	156.49	20.47	0.13
Private 4 Year	36.66	7.70	0.21
Public 2 Year	171.58	23.11	0.13
Private 2 Year	17.48	11.35	0.65
Proprietary 2 Year	19.09	6.63	0.35
Proprietary Less Than 2 Year	35.90	10.88	0.30
Grant Type			
Independent	198.58	21.40	0.11
Dependent	238.63	17.69	0.07
Validation Status			
Non-Validated	180.74	17.37	0.10
Validated	256.47	20.83	0.08
All Categories Combined	437.20	33.17	0.08

TABLE 7-20

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH STUDENT OVERAWARD

Category	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	365.56	31.89	0.09
Private 4 Year	194.35	28.88	0.15
Public 2 Year	122.95	14.86	0.12
Private 2 Year	21.64	8.89	0.41
Proprietary 2 Year	19.73	8.19	0.41
Proprietary Less Than 2 Year	31.88	7.65	0.24
Grant Type			
Independent	183.28	16.82	0.09
Dependent	574.78	23.28	0.04
Validation Status			
Non-Validated	254.70	14.04	0.06
Validated	503.36	28.37	0.06
All Categories Combined	758.07	30.08	0.04

TABLE 7-21

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH STUDENT UNDERAWARD

Category	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	100.10	14.35	0.14
Private 4 Year	62.29	8.83	0.14
Public 2 Year	31.13	6.50	0.21
Private 2 Year	8.03	4.43	0.55
Proprietary 2 Year	6.81	2.46	0.36
Proprietary Less Than 2 Year	8.49	3.18	0.37
Grant Type			
Independent	40.33	6.33	0.16
Dependent	178.48	12.91	0.07
Validation Status			
Non-Validated	64.72	6.47	0.10
Validated	154.09	10.91	0.07
All Categories Combined	218.81	13.47	0.06

TABLE 7-22

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH CASE OVERAWARD

Category	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	499.40	44.10	0.09
Private 4 Year	225.38	31.08	0.14
Public 2 Year	193.05	22.26	0.12
Private 2 Year	24.15	9.11	0.38
Proprietary 2 Year	35.89	19.09	0.53
Proprietary Less Than 2 Year	44.85	6.49	0.14
Grant Type			
Independent	350.98	25.87	0.07
Dependent	671.74	21.92	0.03
Validation Status			
Non-Validated	367.04	21.76	0.06
Validated	655.68	28.26	0.04
All Categories Combined	1,022.72	31.90	0.03

TABLE 7-23

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH CASE UNDERAWARD

Category	Estimate (Thousands)	Standard Error (Thousands)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	189.39	22.08	0.12
Private 4 Year	88.50	12.96	0.15
Public 2 Year	171.97	21.02	0.12
Private 2 Year	18.13	10.87	0.60
Proprietary 2 Year	16.05	5.38	0.33
Proprietary Less Than 2 Year	41.24	11.23	0.27
Grant Type			
Independent	201.20	17.88	0.09
Dependent	324.09	17.60	0.05
Validation Status			
Non-Validated	195.36	14.01	0.07
Validated	329.94	16.30	0.05
All Categories Combined	525.29	24.71	0.05

TABLE 7-24

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN INSTITUTIONAL OVERAWARD

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	531	80	0.15
Private 4 Year	826	89	0.11
Public 2 Year	243	32	0.13
Private 2 Year	595	338	0.57
Proprietary 2 Year	890	251	0.28
Proprietary Less Than 2 Year	681	120	0.18
Grant Type			
Independent	593	63	0.11
Dependent	461	42	0.09
Validation Status			
Non-Validated	537	53	0.10
Validated	528	55	0.10
All Categories Combined	532	45	0.08

TABLE 7-25

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN STUDENT OVERAWARD

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	356	22	0.06
Private 4 Year	325	32	0.10
Public 2 Year	341	24	0.07
Private 2 Year	280	175	0.62
Proprietary 2 Year	617	256	0.41
Proprietary Less Than 2 Year	338	78	0.23
Grant Type			
Independent	568	25	0.04
Dependent	280	11	0.04
Validation Status			
Non-Validated	359	20	0.05
Validated	345	15	0.04
All Categories Combined	350	11	0.03

TABLE 7-26

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN CASE OVERAWARD

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	436	33	0.08
Private 4 Year	479	45	0.09
Public 2 Year	320	25	0.08
Private 2 Year	352	107	0.30
Proprietary 2 Year	850	148	0.17
Proprietary Less Than 2 Year	584	67	0.11
Grant Type			
Independent	620	40	0.06
Dependent	349	16	0.05
Validation Status			
Non-Validated	459	31	0.07
Validated	433	25	0.06
All Categories Combined	442	23	0.05

TABLE 7-27

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN INSTITUTIONAL UNDERAWARD

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	165	17	0.10
Private 4 Year	408	61	0.15
Public 2 Year	233	70	0.30
Private 2 Year	191	59	0.31
Proprietary 2 Year	403	132	0.33
Proprietary Less Than 2 Year	347	109	0.32
Grant Type			
Independent	246	49	0.20
Dependent	232	26	0.11
Validation Status			
Non-Validated	260	39	0.15
Validated	223	31	0.14
All Categories Combined	238	33	0.14

TABLE 7-28

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN STUDENT UNDERAWARD

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	258	21	0.08
Private 4 Year	251	29	0.11
Public 2 Year	257	42	0.16
Private 2 Year	233	181	0.78
Proprietary 2 Year	229	94	0.41
Proprietary Less Than 2 Year	153	45	0.29
Grant Type			
Independent	294	55	0.19
Dependent	238	15	0.06
Validation Status			
Non-Validated	271	41	0.15
Validated	239	13	0.05
All Categories Combined	248	16	0.06

TABLE 7-29

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN CASE UNDERAWARD

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	235	17	0.07
Private 4 Year	286	27	0.09
Public 2 Year	246	60	0.24
Private 2 Year	269	76	0.28
Proprietary 2 Year	430	150	0.35
Proprietary Less Than 2 Year	296	99	0.33
Grant Type			
Independent	252	36	0.14
Dependent	264	18	0.07
Validation Status			
Non-Validated	293	34	0.12
Validated	245	19	0.08
All Categories Combined	259	22	0.09

TABLE 7-30

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN NET INSTITUTIONAL ERROR

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	26	8	0.33
Private 4 Year	14	5	0.38
Public 2 Year	-6	6	1.10
Private 2 Year	-0	1	5.36
Proprietary 2 Year	5	6	1.15
Proprietary Less Than 2 Year	2	3	1.59
Grant Type			
Independent	29	10	0.35
Dependent	11	6	0.53
Validation Status			
Non-Validated	16	6	0.42
Validated	25	9	0.35
All Categories Combined	40	14	0.34

TABLE 7-31

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN ABSOLUTE INSTITUTIONAL ERROR

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	46	9	0.19
Private 4 Year	26	4	0.17
Public 2 Year	26	6	0.24
Private 2 Year	2	1	0.53
Proprietary 2 Year	11	6	0.57
Proprietary Less Than 2 Year	12	3	0.22
Grant Type			
Independent	68	10	0.14
Dependent	55	5	0.09
Validation Status			
Non-Validated	53	6	0.11
Validated	70	8	0.11
All Categories Combined	123	12	0.10

TABLE 7-32

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN NET STUDENT ERROR

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	41	5	0.13
Private 4 Year	19	3	0.18
Public 2 Year	13	2	0.16
Private 2 Year	2	1	0.45
Proprietary 2 Year	4	2	0.47
Proprietary Less Than 2 Year	4	1	0.34
Grant Type			
Independent	36	5	0.13
Dependent	47	4	0.08
Validation Status			
Non-Validated	29	3	0.11
Validated	54	6	0.10
All Categories Combined	83	6	0.07

TABLE 7-33

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN ABSOLUTE STUDENT ERROR

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	62	6	0.09
Private 4 Year	31	4	0.13
Public 2 Year	20	3	0.14
Private 2 Year	3	1	0.38
Proprietary 2 Year	5	2	0.44
Proprietary Less Than 2 Year	5	1	0.31
Grant Type			
Independent	46	4	0.09
Dependent	80	4	0.05
Validation Status			
Non-Validated	43	3	0.06
Validated	83	5	0.07
All Categories Combined	126	6	0.05

TABLE 7-34

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN NET CASE ERROR

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	68	10	0.15
Private 4 Year	33	7	0.21
Public 2 Year	8	6	0.80
Private 2 Year	1	1	0.75
Proprietary 2 Year	9	7	0.72
Proprietary Less Than 2 Year	6	3	0.52
Grant Type			
Independent	66	10	0.15
Dependent	59	6	0.11
Validation Status			
Non-Validated	45	8	0.17
Validated	80	10	0.12
All Categories Combined	125	14	0.11

TABLE 7-35

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
MEAN ABSOLUTE CASE ERROR

Category	Estimate ($)	Standard Error ($)	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	104	12	0.11
Private 4 Year	53	7	0.13
Public 2 Year	41	6	0.15
Private 2 Year	5	2	0.40
Proprietary 2 Year	15	7	0.49
Proprietary Less Than 2 Year	15	3	0.20
Grant Type			
Independent	106	10	0.09
Dependent	127	6	0.04
Validation Status			
Non-Validated	88	6	0.07
Validated	144	8	0.06
All Categories Combined	233	12	0.05

TABLE 7-36

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH UNSATISFACTORY ACADEMIC PROGRESS ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	6,246	3,396	0.54
Private 4 Year	737	697	0.95
Public 2 Year	2,235	1,551	0.69
Private 2 Year	0	0	--
Proprietary 2 Year	0	0	--
Proprietary Less Than 2 Year	0	0	--
Grant Type			
Independent	4,777	2,623	0.55
Dependent	4,441	1,748	0.39
Validation Status			
Non-Validated	784	735	0.94
Validated	8,434	3,597	0.43
All Categories Combined	9,218	3,682	0.40

TABLE 7-37

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH LESS THAN HALF TIME ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	1,466	953	0.65
Private 4 Year	0	0	--
Public 2 Year	0	0	--
Private 2 Year	0	0	--
Proprietary 2 Year	727	676	0.93
Proprietary Less Than 2 Year	784	756	0.96
Grant Type			
Independent	1,466	953	0.65
Dependent	1,511	1,077	0.71
Validation Status			
Non-Validated	1,511	1,077	0.71
Validated	1,466	953	0.65
All Categories Combined	2,977	1,484	0.50

TABLE 7-38

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH INSUFFICIENT PROGRAM LENGTH ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	0	0	--
Private 4 Year	0	0	--
Public 2 Year	0	0	--
Private 2 Year	0	0	--
Proprietary 2 Year	0	0	--
Proprietary Less Than 2 Year	765	726	0.95
Grant Type			
Independent	0	0	--
Dependent	765	726	0.95
Validation Status			
Non-Validated	765	726	0.95
Validated	0	0	--
All Categories Combined	765	726	0.95

TABLE 7-39

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH NONDEGREE PROGRAM ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	0	0	--
Private 4 Year	708	675	0.95
Public 2 Year	0	0	--
Private 2 Year	0	0	--
Proprietary 2 Year	0	0	--
Proprietary Less Than 2 Year	11,184	1,672	0.15
Grant Type			
Independent	8,218	4,890	0.59
Dependent	3,674	3,472	0.95
Validation Status			
Non-Validated	9,690	858	0.09
Validated	2,202	1,181	0.54
All Categories Combined	11,892	1,717	0.14

TABLE 7-40

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH NO STATEMENT OF EDUCATIONAL PURPOSE ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	3,357	1,876	0.56
Private 4 Year	2,186	1,121	0.51
Public 2 Year	2,957	2,226	0.75
Private 2 Year	0	0	--
Proprietary 2 Year	884	883	1.00
Proprietary Less Than 2 Year	0	0	--
Grant Type			
Independent	5,676	2,750	0.48
Dependent	3,708	1,568	0.42
Validation Status			
Non-Validated	2,503	1,442	0.58
Validated	6,881	2,694	0.39
All Categories Combined	9,384	3,451	0.37

TABLE 7-41

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH NO FINANCIAL AID TRANSCRIPT ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	40,779	11,266	0.28
Private 4 Year	22,035	6,025	0.27
Public 2 Year	9,529	4,485	0.47
Private 2 Year	2,186	1,543	0.71
Proprietary 2 Year	6,163	3,407	0.55
Proprietary Less Than 2 Year	1,519	972	0.64
Grant Type			
Independent	46,594	9,018	0.19
Dependent	35,617	5,924	0.17
Validation Status			
Non-Validated	33,110	5,246	0.16
Validated	49,101	8,952	0.18
All Categories Combined	82,211	12,487	0.15

TABLE 7-42

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH LOAN DEFAULT ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	695	664	0.96
Private 4 Year	0	0	--
Public 2 Year	0	0	--
Private 2 Year	0	0	--
Proprietary 2 Year	0	0	--
Proprietary Less Than 2 Year	0	0	--
Grant Type			
Independent	695	664	0.96
Dependent	0	0	--
Validation Status			
Non-Validated	0	0	--
Validated	695	664	0.96
All Categories Combined	695	664	0.96

TABLE 7-43

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH INVALID SAR ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	4,380	1,998	0.46
Private 4 Year	1,464	981	0.67
Public 2 Year	2,566	1,873	0.73
Private 2 Year	0	0	--
Proprietary 2 Year	734	700	0.95
Proprietary Less Than 2 Year	780	765	0.98
Grant Type			
Independent	4,086	2,083	0.51
Dependent	5,837	2,445	0.42
Validation Status			
Non-Validated	2,619	1,950	0.74
Validated	7,304	2,665	0.36
All Categories Combined	9,923	2,918	0.29

TABLE 7-44

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NUMBER OF RECIPIENTS WITH NOT PARENT SCHOOL ELIGIBILITY ERROR

Category	Estimate	Standard Error	Coefficient of Variation
Type and Control of Institution			
Public 4 Year	5,104	2,158	0.42
Private 4 Year	0	0	--
Public 2 Year	0	0	--
Private 2 Year	0	0	--
Proprietary 2 Year	0	0	--
Proprietary Less Than 2 Year	0	0	--
Grant Type			
Independent	1,466	81	0.06
Dependent	3,638	2,150	0.59
Validation Status			
Non-Validated	1,459	1,026	0.70
Validated	3,645	1,241	0.34
All Categories Combined	5,104	2,158	0.42